Small RNAs as Regulators of Early Vertebrate Development

Joanne Harding

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Dr. Caroline S. Hill

A thesis submitted for the degree of

Doctor of Philosophy

University College London

December 2011

Declaration

I, Joanne Harding, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis explores the role of small RNAs as regulators of early vertebrate development using the frog species *Xenopus laevis* and *Xenopus tropicalis* as experimental systems. Firstly, I investigate spatial control of Nodal signalling, which is critical for organisation of the vertebrate body plan during gastrulation, by uncovering the functional relevance and mechanism of spatial regulation of the Nodal coreceptor Cripto-1 in the early *X. laevis* embryo. *Xenopus Cripto-1 (XCR1)* mRNA is ubiquitous in the early embryo, but the protein is absent in the prospective endoderm, from which Nodal signalling originates. Cripto-1 is a stem cell marker and upregulated in many cancers and Nodal signalling is reactivated in melanomas and prostate cancer. Therefore understanding how control of Nodal signalling and XCR1 regulation is achieved may have implications for its misregulation in cancer. I show that spatial regulation of XCR1 is critical for the location, timing and magnitude of Nodal signalling in the early *Xenopus* embryo. Spatial regulation occurs at the level of translational repression, is dependent on the 3' UTR and the microRNA-processing enzyme Dicer and is abolished by mutating four nucleotides in the minimal regulatory region of the 3' UTR. I show that XCR1 is negatively regulated by xla-miR-427 in a 3' UTR-independent mechanism, adding to recent research in *Xenopus* and zebrafish showing microRNA control of Nodal signalling at the ligand, receptor and antagonist level.

Whilst studying *Xenopus* Cripto-1, it became clear that relatively little was known about microRNAs in early vertebrate development. To fill this void, I performed genome-wide sequencing of all small RNAs expressed in the early *X. tropicalis* embryo at three stages of development. This revealed dynamic and localised expression of hundreds of microRNAs in the early embryo and the presence of Piwi-interacting RNA sequences. Extensive validation of the small RNA-seq dataset is presented, and novel microRNAs and a novel class of intronic small RNAs are uncovered within the 95 % of the early vertebrate embryo small RNAome that was previously uncharacterised.

Acknowledgement

I would like to thank my family, friends and colleagues for their support during my studies. Firstly, I thank Caroline Hill for giving me the opportunity to become a graduate student in the Developmental Signalling Laboratory when I was fresh out of university without a great deal of previous research experience, and for supporting and encouraging me throughout my PhD. Thank you to Amanda, Berni, Julie and Laurence for a very warm welcome to the lab. Thanks also to Becky, Debi and Isabel for keeping the lab running smoothly and Eleonora, Marco, Pedro, Eva, Thijs and Tessa for their advice and guidance. I especially want to thank Mary and Sabine for their hard work and support during all the frog experiments we did together. Thanks also to Jane Kirk and Hiro Mahbubani for taking care of our frog colony at Clare Hall.

The small RNA sequencing project would not have been possible without the input of several collaborators. Firstly, thank you to Lyle Zimmerman and lab members at NIMR, in particular Holly Ironfield, for setting up *X. tropicalis* frog experiments to provide me with embryos. Thank you to Eric Miska at the Gurdon Institute for allowing me to join the laboratory for a week to construct the libraries under the supervision of Javier Armisen. Thank you to Stuart Horswell in the CRUK Bioinformatics and Biostatistics group for extensive bioinformatic support and Nik Matthews in the Advanced Sequencing facility for technical expertise. Thanks also to the staff in the LRI service laboratories, in particular the Equipment Park and Light Microscopy facilities.

A special thank you to my parents and my friends especially Ingo, Karen, Kelly, Holly, Sarah, Iain, Tom, Ben, Annabelle (the best landlady in the world) and my brother Dan for all the fun we have had in the last four years and for support when I needed it. To Rob, I know I am very lucky and I feel excited about our future.

Table of Contents

Table of Figures

List of Tables

Abbreviations

ADMP- anti-dorsalising morphogenetic protein

AGO- argonaute protein

ALK4- activin-like receptor kinase-4

APS- ammonium persulfate

AREs- AU-rich elements

BMP- bone morphogenetic protein

BSA- bovine serum albumin

CCR4-Not1- carbon catabolite repression 4-negative on TATA-less

Chip-Seq- chromatin immunoprecipitation sequencing

Chp1- chromodomain protein 1

CPEB- cytoplasmic polyadenylation element binding protein

DAZL- deleted in azoospermia-like

DGCR8- DiGeorge syndrome critical region gene 8

DIG- digoxygenin

DLL1-delta-like 1

DND1- dead end 1

DMSO- dimethyl sulfoxide

4E-BP- mammalian eIF4E binding protein

EDEN- embryonic deadenylation element

EDTA- ethylene diamine tetra-acetic acid

EF1α- elongation factor-1α

EGF-CFC- epidermal growth factor-cripto-FRL1-cryptic

elav- embryonic lethal abnormal visual system

EMT- epithelial to mesenchymal transition

ERF3A- eukaryotic peptide chain release factor 3A

ERK- extracellular signal-regulated protein kinase

ES cell- embryonic stem cell

FGF-R- fibroblast growth factor receptor

GAPDH- glyceraldehyde phosphate dehydrogenase

GPI- glycophosphatidyl inositol

GW182- GW repeat-containing 182 kDa protein

HCG- human chorionic gonadatrophin

HEPES- hydroxyethyl piperazine ethanesulfonic acid

Hh- hedgehog

HIF-1α- hypoxia inducible factor-1α

HP1- heterochromatin protein 1

HuR- Hu antigen R

LB- Luria-Bertani medium

LINE- long interspersed nuclear element

LTR- long tandem repeat

MAPK- mitogen-activated protein kinase

miRNA- microRNA

MMP-1- matrix metalloproteinase-1

MyoD- myogenic differentiation

OAZ- ornithine decarboxylase antizyme 1

ODC- ornithine decarboxylase

PABP- poly(A) binding protein

PAIP- PABP interacting protein

PAM2- PABP-interaction motif 2

PARN- poly(A)-specific ribonuclease

PBS- phosphate buffered saline

piRNA- PIWI-interacting RNA

PIWI- P-element-induced wimpy testis

PNGase- Peptide N-glycosidase F

Pub1- poly(U) binding protein 1

Pum- Pumilio protein

RISC- RNA-induced silencing complex

RITS complex- RNAi-induced transcriptional silencing complex

RNAi- RNA interference

R-Smad- receptor-regulated Smad

SARA- Smad anchor for receptor activation

SDS- sodium dodecyl sulfate

SINE- short interspersed nuclear element

siRNA- small interfering RNA

SSC- saline sodium citrate

SSPE- sodium chloride sodium phosphate- EDTA

TAE- Tris-acetate-EDTA

Taq- *Thermus aquaticus*

TBE- Tris-borate-EDTA

TCF3- transcription factor 3

TDGF1- teratocarcinoma-derived growth factor 1

tdrd7- tudor domain containing protein 7

TE- Tris-EDTA

TEMED- tetramethylethylenediamine

TGF-β- transforming growth factor-β

TNF-α -tumour necrosis factor-α

TTP- tristetrapolin

U6- spliceosomal small RNA U6

uPA- urokinase plasminogen activator

3’ UTR- 3’ untranslated region

VEGF- vascular endothelial growth factor

XCR1- *Xenopus* Cripto-1

Xnr- *Xenopus* Nodal-related

ZEB1- zinc finger E-box binding homeobox 1

Chapter 1. Introduction

The initial aim of this thesis was to understand the mechanism of spatial regulation of *Xenopus* Cripto-1 (XCR1), a coreceptor for the TGF-β superfamily member, Nodal. This led to further investigation of potential regulators of XCR1 that can modulate gene expression via the 3' UTR including microRNAs and RNA-binding proteins. In this chapter I will introduce the *Xenopus* embryo as a model for understanding signalling events that control early vertebrate development and explore the roles that regulators such as ligands of the TGF-β superfamily, Cripto co-receptors, small RNAs and RNA-binding proteins play to regulate gene expression in early development.

1.1 Spatial regulation of transcription and protein localisation specify and pattern the germ layers of the early *Xenopus* embryo

During gastrulation, the embryo changes from a ball of cells to a structure containing three layers by morphogenetic cell movements. The three 'germ layers' are the endoderm, mesoderm and ectoderm. The innermost layer is the endoderm, which becomes the epithelial lining of the gut and lungs and organs such as the liver and pancreas. Mesodermal derivatives form from the middle layer and include the notochord, blood, somites, head mesoderm, the heart, blood vessels and limbs. The skin and nervous system form from the outermost layer, the ectoderm (Moody, 1987). Germ layer formation and patterning in *Xenopus* requires signalling events along two embryonic axes: the animal-vegetal axis and the dorsal-ventral axis (Figure 1.1).

The animal-vegetal axis is determined by the structure of the egg, which consists of a pigmented animal hemisphere and a yolky vegetal hemisphere. In contrast, the dorsal-ventral axis is established at fertilisation. The point of sperm entry on the egg defines the future location of the dorsal lip of the blastopore directly opposite the point of sperm entry (Gerhart et al., 1989), and is where the prospective mesoderm and endoderm will start to involute during gastrulation. Transcription from the newly formed zygotic genome does not begin until the mid-blastula transition (embryonic stage 9) and maternal mRNAs and proteins are expressed up until this point (Newport and

Kirschner, 1982a, b). Transcriptional activation at the mid-blastula transition in *Xenopus* and zebrafish correlates with the acquisition of the activatory trimethylated H3K4 histone mark (Akkers et al., 2009; Vastenhouw et al., 2010). Maternal mRNAs are cleared at the mid-blastula transition by miR-427 in *Xenopus* and in zebrafish, by the corresponding homologue, miR-430 (Lund et al., 2009; Giraldez et al., 2006).

Sperm entry also initiates rotation of the vegetal cortex (a layer of actin filaments) away from the sperm entry point and towards the future dorsal region by 30°(Schroeder and Gard, 1992). Cortical rotation requires microtubule polymerisation and is blocked by UV light or microtubule polymerisation inhibitors (Elinson and Rowning, 1988; Scharf and Gerhart, 1983). Cortical rotation leads to the accumulation of the transcriptional cofactor β-catenin in the future dorsal region of the embryo (Rowning et al., 1997). Nuclear β-catenin interacts with the transcription factor Tcf3 to regulate target gene expression (Moon and Kimelman, 1998). mRNA of the transcription factor VegT is vegetally localised in the oocyte, leading to vegetal expression of maternal VegT protein, which induces expression of *Xenopus* Nodal-related TGF-β ligands (Xnrs)(Zhang and King, 1996). The overlapping zone of dorsal β-catenin and vegetal VegT expression forms the Nieuwkoop centre, which initiates expression of Xnr5 and Xnr6 when zygotic transcription begins (Takahashi et al., 2000). Expression of Xnr1, 2, and 4 is also enriched dorsally (Agius et al., 2000). This is assumed to create a gradient from high Nodal ligand levels in the vegetal dorsal region to low Nodal ligand levels on the opposing ventral side of the embryo.

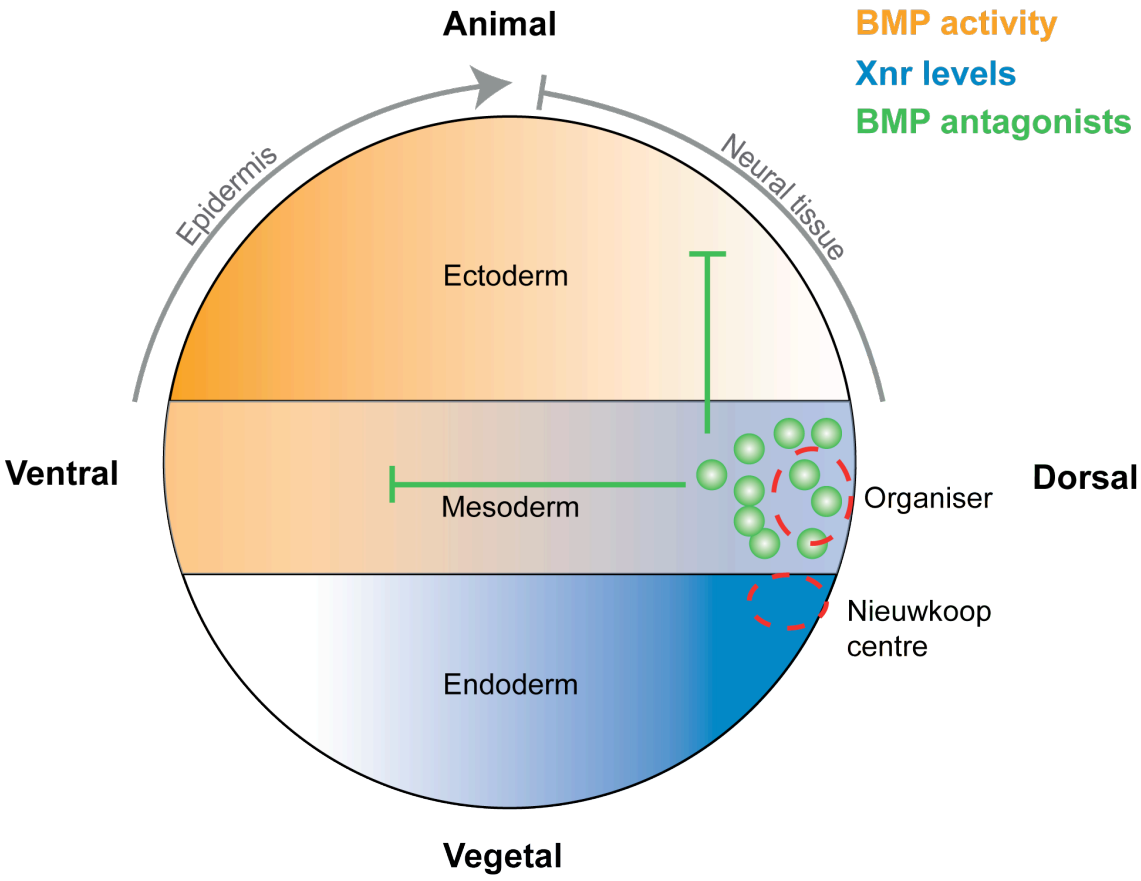


Figure 1.1 Germ layer specification and patterning requires spatial control of TGF-β superfamily signalling across the animal-vegetal and dorsal-ventral embryonic axes in the early *Xenopus* embryo

Schematic lateral view of a stage 10 *Xenopus* embryo showing the prospective endoderm, mesoderm and ectoderm. Vegetal Xnr signalling is shown in blue, and the gradient indicates the dorsal-ventral concentration gradient of Xnrs established by the Nieuwkoop centre. Xnrs are also expressed in mesoderm. The Nieuwkoop centre induces the Spemann organiser, which secretes the BMP antagonists chordin, noggin and follistatin (shown in green). These inhibitors are diffusible and act over a long range to establish a gradient of BMP ligand activity that is higher on the ventral side than the dorsal side (orange gradient). High ventral BMP activity in the prospective ectoderm specifies an epidermal cell fate, whereas low BMP activity on the dorsal side specifies neural tissue (shown by grey arrows).

Nodal ligands are required for patterning mesoderm and are transduced to the nucleus by phosphorylated Smad2 (Figure 1.2 and section 1.1.2 for a detailed discussion of the pathway). If Nodal ligands are bound by an artificial extracellular inhibitor, Cerberus-short (Piccolo et al., 1999), mesoderm specification is blocked and endodermal markers are downregulated (Engleka et al., 2001; Piccolo et al., 1999). An activity gradient of Nodal ligands that is high on the dorsal side and lower on the ventral side is thought to lead to a variety of endodermal cell fates along the dorsal-ventral axis in the prospective endoderm (Figure 1.1), although it is not known whether a linear dose-response relationship is obeyed as reported for the TGF-β ligand Activin (Gurdon et al., 1994). Activation of the transcription factor sox17α by vegetal VegT is thought to prevent mesodermal gene expression in the prospective endoderm in response to Nodal signals (Engleka et al., 2001).

The Nieuwkoop centre establishes the Spemann Organiser above the dorsal lip of the blastopore by secretion of a diffusible signal thought to be Wnt11 (Tao et al., 2005). When the Spemann Organiser is excised and grafted onto the opposing ventral side of another embryo, a secondary body axis forms, demonstrating the powerful inductive capacity of the region (Spemann and Mangold, 2001). The homeobox transcription factor goosecoid is expressed in the organiser region and ventral ectopic expression of *goosecoid* mRNA has the same effect as an organiser graft, suggesting that goosecoid plays a major role in controlling gene expression in the organiser (Cho et al., 1991).

The organiser region secretes antagonists of the TGF-β superfamily ligands, BMPs (bone morphogenetic proteins). In *Xenopus*, BMP2, BMP4 and BMP7 pattern ectoderm and mesoderm and are expressed in ectoderm and mesoderm, with BMP4 expression higher on the ventral side (Hemmati-Brivanlou and Thomsen, 1995). The mechanism of BMP signal transduction to the nucleus is described in Figure 1.2. Secretion of the BMP inhibitors chordin, noggin and follistatin from the organiser in the dorsal marginal zone antagonises the ventralising activity of BMPs to establish a gradient of BMP ligand activity (not concentration) along the dorsal-ventral axis that is high on the ventral side and low on the dorsal side (Figure 1.1)(Piccolo et al., 1996; Zimmerman et al., 1996). This can be directly observed by immunostaining of phospho-Smad1, the intracellular transducer of BMP activity (Faure et al., 2000; Wu et al., 2011). Low BMP activity on

the dorsal side of the prospective ectoderm specifies neural tissue, whereas high BMP activity specifies epidermal tissue (Wilson and Hemmati-Brivanlou, 1995). A range of mesoderm types form along the dorso-ventral BMP gradient (Graff, 1997). Dorsal mesoderm derivatives form where BMP activity is low and include head mesoderm and notochord. Ventral mesoderm derivatives form where BMP activity is high and include muscle and kidney.

The ventralising activity of BMPs is observed when a late blastula embryo is dissected along the dorsal-ventral axis. The dorsal half forms a complete, but smaller embryo, whereas the ventral half forms a 'belly piece' lacking any dorsal structures (Reversade and De Robertis, 2005; Spemann, 1938). The dorsally expressed BMP, ADMP, is thought to allow the dorsal half to recover a BMP activity gradient to correctly pattern the embryo (Reversade and De Robertis, 2005). The antagonism of BMP activity by the organiser is subject to a further layer of regulation by the opposing activities of the dorsalising organiser gene goosecoid, and the ventralising BMP target gene vent1, which downregulate each other by mutual transcriptional repression (Sander et al., 2007).

In short, spatial control of transcription and protein localisation along the animal-vegetal and dorsal-ventral axes are required for germ layer formation and patterning in the early *Xenopus* embryo. For example, maternal *VegT* mRNA localisation, dorsal localisation of β-catenin, dorsal *Xnr* transcription from the Nieuwkoop centre to establish the Nodal activity gradient and the Spemann organiser, and subsequent establishment of the BMP activity gradient are critical events requiring spatial regulation.

1.1.1 Parallels with other vertebrates

Spatial control of Nodal signalling along embryonic axes is required for germ layer formation in other vertebrate systems including zebrafish, chick and mouse (reviewed in (Wu and Hill, 2009; Zorn and Wells, 2009). In general, endoderm is derived from Nodal-producing cells leading to mesoderm induction in adjacent cells. In zebrafish, signalling events across the dorsal-ventral axis are similar to those in *Xenopus*, with a dorsal organiser region known as the shield establishing the dorsal-ventral BMP activity

gradient. The zebrafish Nodal ligands are *squint* and *cyclops*, and are produced in a marginal layer of cells closest to the yolk in the zebrafish epiblast. During gastrulation, marginal cells involute under the epiblast to form the germ layers.

In contrast to *Xenopus* and zebrafish, there is one Nodal gene in chick and mouse. Again, Nodal signalling is critical for germ layer formation, but is spatially regulated across the anterior-posterior axis. Nodal activity is higher in the posterior epiblast cells due to antagonism by the inhibitors *lefty* and *cerberus* in the anterior cells (see Figure 1.2). High Nodal activity initiates gastrulation by forming the primitive streak, which is where posterior epiblast cells involute to form the germ layers.

1.1.2 TGF-β signal transduction in early *Xenopus* development

As illustrated in Figure 1.2, there are traditionally two main branches of TGF-β signalling in *Xenopus* development, the BMP pathway, and the Activin-related pathway. Signalling by the BMP ligands BMP2, 4 and 7 through a Type I and Type II receptor complex (ALK3/ALK2 and BMPRII respectively) leads to phosphorylation and activation of Smad1 and Smad5 in *Xenopus*. The activated Smad complexes accumulate in the nucleus and bind DNA in association with transcription factors to activate BMP target genes. The phospho-Smad1 gradient reflects the dorsal-ventral BMP activity gradient, and is established at late stage 9 (Faure et al., 2000).

Activin-related pathway ligands include Activin, the *Xenopus* Nodal ligands Xnr1, 2, 3, 4, 5 and 6, Derriere and Vg-1. With the exception of Xnr3, a neural rather than a mesoderm inducer (Hansen et al., 1997), Activin-related pathway ligands induce phosphorylation of Smad2 and Smad3 via a Type I-Type II receptor complex (ALK4 and ActRIIA/ActRIIB) and are crucial for mesoderm formation. ALK7 can also act as a Type I receptor for Nodal ligands (Reissmann et al., 2001). There is very little Smad3 in early *Xenopus* development (Howell et al., 2001), so phospho-Smad2 rather than phospho-Smad3 is the main readout of TGF-β signalling by activin-related ligands. Phospho-Smad2 expression begins at stage 9, after the mid-blastula transition and mirrors the Nodal activity gradient. Phospho-Smad2 is dorsally localised at stage 9, and found in the vegetal pole and not the animal pole at stage 10, with some phospho-Smad2 in the marginal zone (Faure et al., 2000). A Smad2 isoform lacking the amino

acids encoded by exon3, Smad2Δexon3, is also responsive to activin-related ligands and mirrors the expression of full length Smad2 (Faure et al., 2000).

In addition to their role as potent mesoderm inducers (Section 1.1), asymmetrical expression of Nodal ligands controls left-right asymmetry later in development, for example in organogenesis. Strikingly, the direction of a snail shell coil can be reversed by altering the expression domain of Nodal, and loss of shell coiling is observed when Nodal is inhibited (Grande and Patel, 2009; Kuroda et al., 2009).

Recent work indicates that the simple division of TGF-β signalling into the BMP pathway signalling via Smad1/5 (and Smad 8 in humans) and the Activin-related pathway signalling through Smad2/3 is not entirely accurate. Phosphorylation of Smad1 and Smad5 in response to TGF-β stimulation (equivalent to Activin/Nodal stimulation in *Xenopus*) has been observed in multiple human cell lines and leads to the formation of novel Smad complexes, such as phosphorylated Smad1-Smad2 complexes instead of the traditional phospho-Smad2/3-Smad4 complex (Daly et al., 2008) (Liu et al., 2009). However, Smad1 and Smad5 phosphorylation in response to Activin-related ligands has not yet been demonstrated in *Xenopus*.

Moreover, TGF-β ligands from both branches can heterodimerise and signal. In addition to endogenous BMP2-BMP7 heterodimer activation of BMP signalling in zebrafish (Little and Mullins, 2009), BMP4 can heterodimerise with the Activin-related ligand Derriere when overexpressed in *Xenopus* (Eimon and Harland, 2002). Nodal and BMP7 form dimers when immature ligands lacking cleavage sites are overexpressed in *Xenopus*, and coexpression of mature ectopic BMP7 and Nodal inhibits Smad2 and Smad1 phosphorylation, suggesting that BMP7-Nodal heterodimer formation could be responsible for the mutual antagonism of BMP and Nodal signalling (Yeo and Whitman, 2001). The Vg-1-related ligand, GDF1, can heterodimerise with Nodal in the mouse to form a more potent activator of the Nodal pathway than either GDF1 or Nodal homodimers (Tanaka et al., 2007).

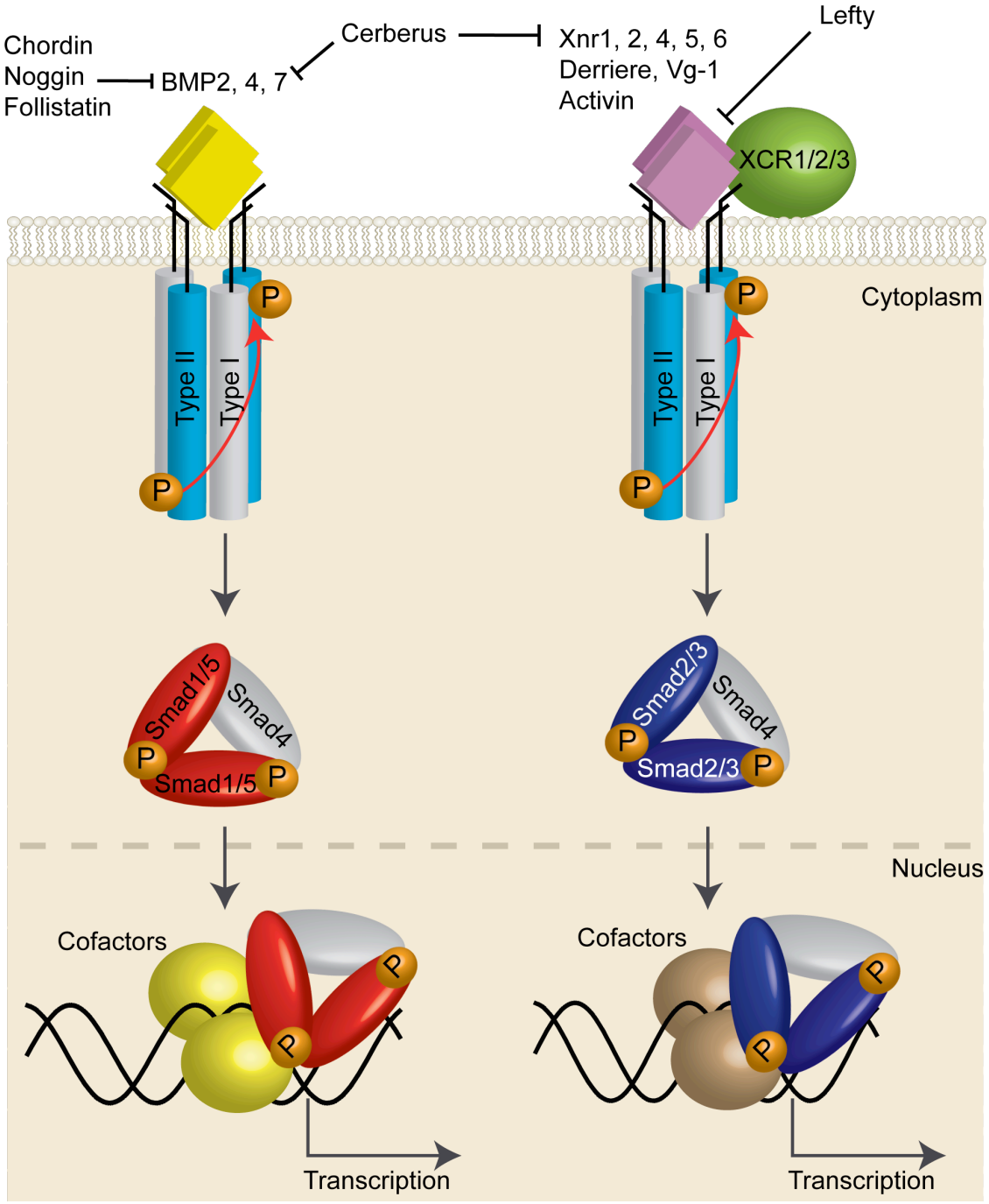


Figure 1.2 Schematic of TGF-β superfamily signal transduction in *Xenopus*
Traditionally, there are two branches of TGF-β signalling in *Xenopus*: the BMP pathway (ligands are BMP2, 4, 7) and the Activin-related pathway (Xnrs, Derriere, Vg-1 and Activin). TGF-β ligands induce complex formation between the Type II and Type I receptors, leading to phosphorylation and activation of the Type I receptor serine-threonine kinase activity, which phosphorylates R-Smads. R-Smads form complexes with the co-Smad, Smad4 and accumulate in the nucleus where they bind DNA in association with cofactors to initiate gene regulation. Antagonism by extracellular inhibitors is indicated and the coreceptors XCR1, 2 and 3 are critical for signal transduction by Xnr ligands of the Activin-related pathway.

However, there are some clear differences in regulation of the BMP and Activin-related branches of TGF-β signalling in *Xenopus* (Figure 1.2). Firstly, signal transduction by Nodal ligands requires coreceptors of the EGF-CFC family (Yeo and Whitman, 2001), which are extracellular membrane-linked proteins that contain a conserved EGF-like domain and a CFC domain (see section 1.2). There are three EGF-CFCs in *Xenopus*, XCR1, 2 and 3 (Dorey and Hill, 2006). Whereas Xnrs require EGF-CFCs to signal, other ligands of the Activin-related pathway, such as Activin, do not. Moreover, EGF-CFCs are sensitive to antagonism by the diffusible inhibitor Lefty (Meno et al., 1999), which also binds Nodals directly to decrease ligand availability for receptor binding (Chen and Shen, 2004). In contrast, the extracellular inhibitor Cerberus inhibits both the BMP and Activin-related branches by direct binding to ligands to decrease ligand availability for receptor binding (Piccolo et al., 1999). Cerberus is expressed in the anterior endoderm, represses trunk and tail mesoderm formation and is important for head induction, as targeted overexpression of Cerberus leads to the formation of an extra head (Bouwmeester et al., 1996).

Moreover, transcriptional responses specific to BMP or Activin-related ligand stimulation are achieved by the interaction of Smads with particular transcriptional cofactors. Smads cannot activate transcription from DNA alone and instead require a chromatin template (Ross et al., 2006). Smad2-Smad4 complexes are recruited to promoters by interaction with FoxH1 or Mix family transcription factors containing a Smad interaction motif (SIM) (Germain et al., 2000), or p53 (Cordenonsi et al., 2003), leading to transcription from Activin/Nodal target genes. For example, Xnr1 transcription is activated by Smad2-Smad4 and FoxH1, leading to upregulation and maintenance of Nodal signalling (Osada et al., 2000). The transcription factor OAZ is thought to recruit Smad1-Smad4 complexes to DNA to activate transcription of the BMP target gene, Vent2, although the OAZ recognition motif is similar to a Smad-binding element identified in *Drosophila* (Hata et al., 2000) (von Bubnoff et al., 2005).

Recent work has uncovered microRNA regulation of Activin-related pathway signalling at multiple levels. In zebrafish, the Nodal ligand *squint* and the inhibitor *lefty* are repressed by miR-430 (Choi et al., 2007). Similarly, *lefty* is a target of the miR-430

homologue in *Xenopus*, miR-427, which also targets the Nodal ligands Xnr5 and Xnr6 (Rosa et al., 2009). The human homologue, miR-302 targets *lefty* (but not Nodal ligands), thereby enhancing Nodal signalling. At the receptor level, the Type II receptor ActRIIA is targeted by miR-15 and miR-16 to downregulate Nodal signalling and restrict the size of the organiser in *Xenopus* (Martello et al., 2007). miR-15 and miR-16 expression is thought to be ventrally enriched, with the ActRIIA protein ventrally localised. Intriguingly, the signal thought to establish the ventral expression of miR-15 and miR-16 is suggested to be inhibition of miR-15 and miR-16 expression by dorsal β-catenin (established by cortical rotation, see section 1.1) (Martello et al., 2007).

1.2 Role of the Nodal coreceptor, Cripto-1

Cripto was first identified in humans in undifferentiated teratocarcinoma cells (hence Cripto is also known as teratocarcinoma-derived growth factor 1, TDGF1), and in *Xenopus* as a secreted ligand activator of FGF receptor signalling, FRL1 (for FGF receptor related ligand)(Ciccodicola et al., 1989; Kinoshita et al., 1995). EGF-CFCs were first identified as essential for Nodal signalling by genetic studies in zebrafish. Maternal-zygotic mutants for the zebrafish EGF-CFC, *oep* (one-eyed pinhead), do not respond to Nodal ligands and phenocopy mutants of the Nodal ligands *cyclops* and *squint* (Gritsman et al., 1999). EGF-CFCs are conserved in vertebrates and other family members include human Cripto, mouse Cripto and Cryptic and *Xenopus* Cripto-1, 2 and 3 (XCR1, 2 and 3) (Dorey and Hill, 2006; Onuma et al., 2006a; Saloman et al., 2000). Figure 1.3 shows the conserved EGF-like and CFC domains in zebrafish, frog and human EGF-CFCs. EGF-CFCs also contain an N-terminal signal sequence, are linked to the cell membrane by a glycophosphatidyl inositol (GPI) tag and are modified by *N*-glycosylation, including a fucosylation modification that is essential for Nodal signalling in mammalian cells (Schiffer et al., 2001). The evidence for the role of EGF-CFCs as coreceptors for Nodal is the interaction of Cripto with ALK4 via the CFC domain in *Xenopus*, which is essential for Nodal binding the Type I-Type II receptor complex and Smad2 activation (Yeo and Whitman, 2001). Moreover, Cripto can bind Nodal directly in mammalian cells and Xnr1 in *Xenopus* and enhances Nodal signal transduction via ALK7 (Reissmann et al., 2001; Yan et al., 2002). Transgenic mice

lacking Cripto-1 are embryonically lethal because they are unable to gastrulate, illustrating the importance of Cripto-1 coreceptor activity (Ding et al., 1998).

Figure 1.3 EGF-CFC family members are conserved in vertebrates

Protein alignment of *X. laevis, X. tropicalis*, human and zebrafish EGF-CFCs coloured for percentage identity in Jalview and generated in Clustal W. *Xenopus laevis* Cripto-1, 2 and 3 (XCR1, 2 and 3) and the proteins resulting from the XCR1 pseudoalleles, XCR1α and XCR1β are shown. The founding EGF-CFC family member is zebrafish *one-eyed pinhead* (oep). The conserved EGF-like domain and CFC domain are highlighted in red. The CFC domain is named after founding members of the protein family Cripto, FRL-1 (later renamed XCR1) and Cryptic.

In *Xenopus*, Cripto-1, 2 and 3 (XCR1, 2 and 3) are differentially expressed in early development. *XCR1* and *XCR3* mRNA is expressed maternally and zygotically until neurulation begins, whereas *XCR2* is only expressed after gastrulation and persists to tadpole stage (Dorey and Hill, 2006; Onuma et al., 2006b). Whereas XCR2 is important for left-right patterning, XCR1 and 3 are essential for Nodal signalling in *Xenopus* (Dorey and Hill, 2006). The initial characterisation of XCR1 an FGF receptor ligand instead of a Nodal coreceptor could be explained by ERK induction due to the wound-healing response when animal caps are cut and/or the neural phenotype of XCR1 morphants (Christen and Slack, 1999) (Yabe et al., 2003). Moreover the EGF-like domain of Cripto-1 can activate MAP kinase in mammalian cells (Kannan et al., 1997). However, the critical role of XCR1 and XCR3 in *Xenopus* embryos is to act as coreceptors for Nodal ligands (Dorey and Hill, 2006).

XCR3 is expressed ubiquitously in the embryo, but XCR1 protein expression is temporally and spatially regulated (Dorey and Hill, 2006). *XCR1* mRNA levels are constant whilst XCR1 protein peaks at gastrulation. This post-transcriptional regulation has been shown to be due to the increased association of XCR1 mRNA with polyribosomes and elongation of the mRNA poly(A) tail after fertilisation, which promotes translation of *XCR1* mRNA (Zhang et al., 2009). *XCR1* mRNA is ubiquitously expressed at gastrulation (stage 10 of *Xenopus* development); however, XCR1 protein is only found in the animal cap and marginal zone. XCR3 is therefore the only Cripto expressed in the vegetal pole where Nodal ligands are synthesised (see section 1.1).

XCR1 and XCR3 synergise with different Xnrs in early *Xenopus* development to cause downstream phosphorylation and activation of Smad2 (Dorey and Hill, 2006). XCR1 synergises with Xnr1, 2, 5 and 6, and XCR3 synergises with Xnr 1, 2, 4, 6 and Derriere. Since XCR3 alone is expressed in the vegetal pole at gastrulation, only XCR3-dependent ligands are able to signal. In contrast, since Xnr5 synergises with XCR1 and not XCR3, Xnr5 signalling is potentiated in the marginal zone and not in the vegetal pole. In this way, spatial regulation of XCR1 expression controls the spatial activation of the Nodal pathway. The mechanism of spatial regulation of XCR1 has yet to be elucidated, but involves spatial control of translation in the gastrula-stage embryo and is

a major point of investigation in this thesis (see Chapter 3). One study has also suggested that XCR1 interacts with maternal Wnt11 and acts as a coreceptor for Wnt signalling (Tao et al., 2005), however this result has not been reproduced in any other systems and there is very little maternal XCR1 protein (Dorey and Hill, 2006).

In cell culture, Cripto acts as a secreted growth factor (Brandt et al., 1994). It is thought that Cripto can act as a soluble ligand and a membrane-bound coreceptor by its release from the cell membrane. Both forms of Cripto can activate Nodal signalling, and a C-terminally truncated *oep* mRNA lacking the GPI-tag sequence can rescue the phenotype and Nodal signalling activity of *oep* mutants (Yan et al., 2002; Zhang et al., 1998). In contrast to the activatory role of Cripto as a Nodal coreceptor in developmental systems, Cripto has been shown to bind TGF-β via the EGF-like domain to antagonise TGF-β signalling in multiple human cell lines and thereby inhibit the growth suppressive effects of TGF-β in mammary epithelial cells (Gray et al., 2006).

Cripto is also thought to regulate Nodal signalling at the level of ligand processing and trafficking. In cell culture, Cripto interacts with the pro-domain of the Nodal ligand precursor and with the convertases Furin and PACE4 at the cell surface to form a ligand-processing complex (Blanchet et al., 2008a). Cripto also promotes the internalisation of Nodal-receptor complexes to endosomes for signalling and Smad activation via interaction with SARA (Smad anchor for receptor activation), rather than lysosomal degradation (Blanchet et al., 2008a; Blanchet et al., 2008b).

1.3 Cripto-1 in cancer

Cripto-1 is implicated in tumourigenesis due to its overexpression in many cancers including breast cancer, colorectal tumours, pancreatic cancer, gastric cancer and oral squamous cell carcinoma (Salomon et al., 1999) (Ciardiello et al., 1991) (Friess et al., 1994) (Kuniyasu et al., 1991) (Yoon et al., 2011). Cell culture studies have shown that Cripto-1 expression promotes transformation, proliferation, anchorage-independent growth, cell migration, angiogenesis and inhibits differentiation of mammary cells (Ciccodicola et al., 1989) (Adamson et al., 2002). Transgenic mice overexpressing human Cripto-1 in the mammary gland develop hyperplasias and adenocarcinomas and show upregulation of EMT markers, demonstrating the *in vivo* role of Cripto-1 in tumourigenesis (Strizzi et al., 2004; Wechselberger et al., 2005) (Bianco et al., 2005). Therefore, understanding how Cripto-1 is regulated in *Xenopus* may have implications for the mechanism of Cripto-1 deregulation in cancer.

Cripto-1 protein in the plasma is a biomarker for breast and colon cancer detection, and antibodies targeting the CFC domain successfully inhibit tumour cell growth in xenograft models of testicular and colon cancer (Bianco et al., 2006) (Adkins et al., 2003). Cripto-1 is also a marker for undifferentiated human embryonic stem cells (Brivanlou et al., 2003). Within a human embryonal carcinoma stem cell population (stem cells from teratocarcinomas), high Cripto-1 expression correlates with tumourigenesis and the expression of pluripotent embryonic stem cell marker genes such as Oct4 and Nanog (Watanabe et al., 2010). Similarly, melanomas secrete Nodal and Nodal is considered a biomarker for melanoma progression (Topczewska et al., 2006). However, very little Cripto-1 expression was detected in melanomas (Strizzi et al., 2009). Instead, Nodal expression in melanomas appears to be dependent on Notch4 (Hardy et al., 2010). Strikingly, transplantation of melanoma cells onto the zebrafish embryo induces secondary axis formation via upregulation of axial mesoderm through Nodal secretion (Topczewska et al., 2006). These observations provide evidence for the view that the reactivation of embryonic signalling pathways is important in tumourigenesis.

1.4 Overview of small RNA biogenesis and modes of action

Whereas ligands of the TGF-β superfamily regulate gene expression by initiating specific transcriptional programs in early development, small RNAs regulate gene expression at the level of post-transcriptional regulation or epigenetic mechanisms. Based on biogenesis criteria, there are three main classes of small RNAs: siRNAs, microRNAs and piRNAs. Small RNA biogenesis requires proteins of the Argonaute (Ago) family (comprised of Ago and Piwi proteins), and the RNAseIII enzymes Drosha and Dicer (reviewed in (Kim et al., 2009)).

1.4.1 siRNA biogenesis and mechanism of action

Small interfering RNAs (siRNAs) are ~21-24 nt small RNAs that have antisense complementarity to target mRNA and silence gene expression by causing mRNA degradation. siRNAs were first identified in response to viral infection in plants and have since been sequenced in *Drosophila*, *C. elegans*, mouse oocytes and embryonic stem cells and most recently in *X. tropicalis* (Czech et al., 2008) (Hamilton and Baulcombe, 1999) (Ambros et al., 2003b) (Armisen et al., 2009; Babiarz et al., 2008; Watanabe et al., 2008).

Endogenous siRNAs are generated from long double-stranded RNA precursors that are transcribed from repetitive loci in the genome such as transposons or siRNA clusters (Figure 1.4). Double-stranded endo-siRNA precursors form from sense and antisense transcripts from the same or different loci, or from self-complementary transcripts that fold into hairpin structures. Double-stranded RNA can also be introduced experimentally by microinjection or transfection (exogenous dsRNA). Double stranded RNA from endogenous or exogenous sources is processed by Dicer into short 21-24 nt dsRNA duplexes. One strand of the siRNA duplex is then degraded and the other interacts with Ago2 (a member of the Ago subfamily with 'slicer' endonuclease activity) to guide interaction of Ago2 with the target mRNA to cause mRNA degradation (Figure 1.4). Interaction of endo-siRNAs with the RNA-binding protein R2D2 is necessary to restrict endo-siRNA loading to Ago2, rather than Ago1 in *Drosophila* (Okamura et al., 2011). In plants, yeast and *C. elegans*, an RNA-dependent RNA polymerase can synthesise double-stranded endo-siRNA precursors from a single-

stranded RNA template to cause gene silencing by RNA-interference (RNAi). Both endo-siRNA and exogenous siRNA (exo-siRNA) dsRNA precursors can be synthesised by RNA-dependent RNA polymerase to amplify the RNAi effect in these systems.

Experimental siRNA inhibition of gene expression or RNAi is effective in many organisms, but not in *Xenopus* or zebrafish. Morpholino oligonucleotides have been used as alternative tools to siRNAs for loss of function studies. A recent study showed that Ago proteins are saturated with siRNAs in the *Xenopus* embryo before the onset of zygotic Ago expression at the mid-blastula transition (Lund et al., 2011). Lack of Ago protein availability inhibits double-stranded siRNA processing by Dicer into single-stranded small RNAs (see Figure 1.5). Overexpression of human Ago proteins renders the embryo competent for siRNA-mediated knockdown of gene expression (Lund et al., 2011).

In addition to post-transcriptional gene silencing by degradation of target transcripts, siRNAs can silence gene expression at the level of chromatin in plants, yeast, and animals. In plants, siRNAs bound to Ago4 can direct DNA methylation and acquisition of repressive H3K9 methylation at target repeat sequences in the genome. Often, siRNAs are often complementary to a repeat in the promoter of a silenced gene (Verdel et al., 2009). In yeast, a complex containing siRNA, Ago1 and Chp1 (which binds trimethylated H3K9) directly binds DNA to mediate siRNA-directed heterochromatin formation (Verdel et al., 2004).

Exogenous siRNAs complementary to the promoter DNA can modify chromatin in mammalian cells, involving Ago1 or Ago2 recruitment to the promoter (Janowski et al., 2006). Examples of exogenous siRNA-directed chromatin modification include repression of the EF1A locus in humans, requiring H3K27 methylation (Weinberg et al., 2006) and repression of the INK4/ARF locus (Gonzalez et al., 2006). On the other hand, exogenous 21 nt dsRNAs complementary to the promoters of E-cadherin, p21 and VEGF caused loss of repressive H3K9 methylation and Ago2-dependent activation of gene expression (Li et al., 2006). The mechanism for exogenous siRNA-mediated activation of gene expression is unknown, but was not effective for dsRNAs targeted to CpG-rich regions of the same promoters. One study has shown that transcription

overlapping the promoter is a feature of exogenous siRNA-directed chromatin modification of the INK4/ARF locus, suggesting that RNA:RNA interactions at the promoter may be critical, in contrast to previous studies where increased methylation of promoter DNA was observed (Gonzalez et al., 2008) (Morris et al., 2004).

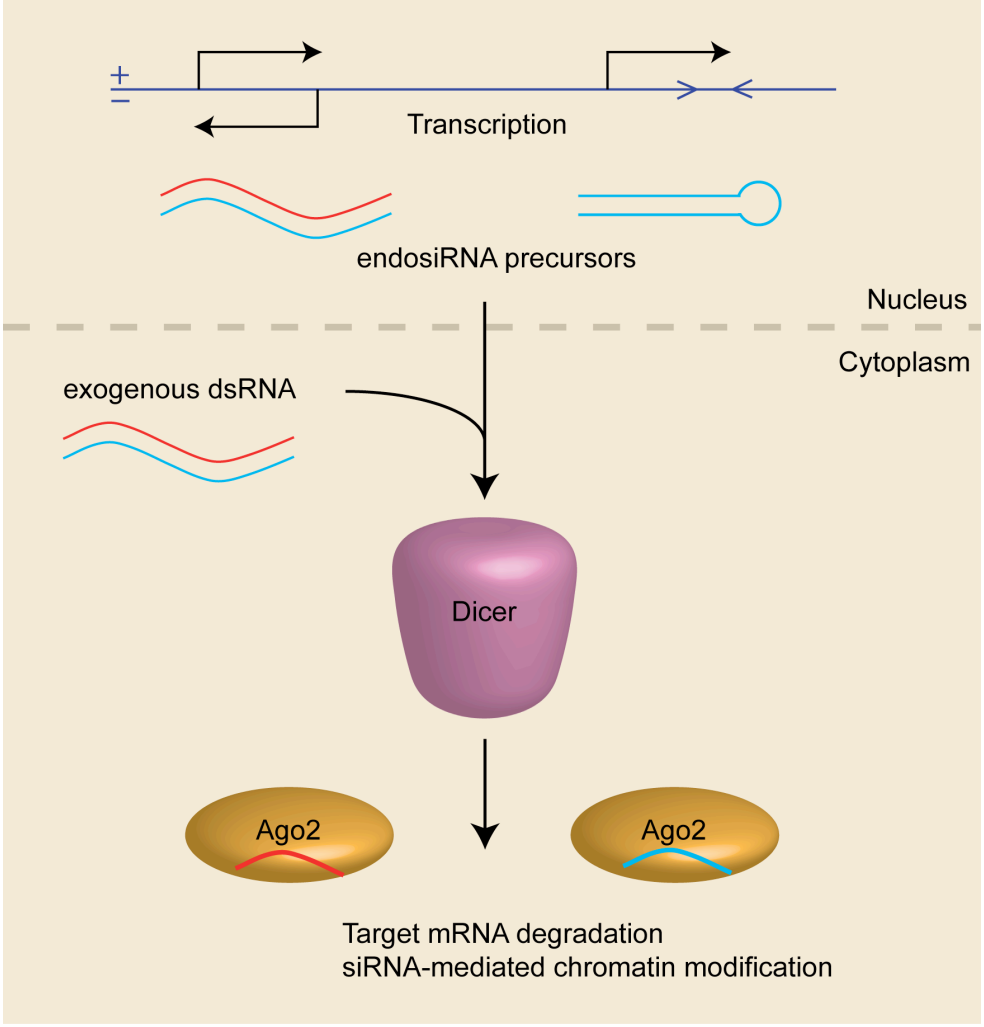


Figure 1.4 siRNA biogenesis pathway

Endogenous siRNAs (endo-siRNAs) originate from transcription from siRNA loci in the genome, which may be transposons, other repetitive elements or siRNA clusters. Double-stranded RNA endo-siRNAs precursors can be derived from overlapping sense and antisense transcripts or hairpin structures. Exogenous dsRNA from viruses or experimentally introduced dsRNA and endo-siRNA precursors are processed by Dicer in the cytoplasm to single-stranded short interfering RNAs (siRNAs) which are bound by Argonaute-2 proteins (Ago2). Degradation of the complementary target mRNA by Ago2 slicer activity, or siRNA-mediated remodelling of chromatin for epigenetic silencing may then occur.

1.4.2 MicroRNA biogenesis and mechanism of action

MicroRNAs are 22-23 nt small RNAs that are present in many species, ranging in complexity from single-celled algae to humans. Currently, there are 16,772 microRNAs listed in the miRbase registry (miRbase 17), of which 1424 are human microRNAs and 208 are *Xenopus* miRNAs (Griffiths-Jones et al., 2006). MicroRNAs are well conserved in bilateran animals. For example, 54 % of *Drosophila* miRNAs have a human homologue (Ibanez-Ventoso et al., 2008). MicroRNAs generally bind the 3'UTR of target transcripts to negatively regulate gene expression post-transcription. MicroRNA loci are often found in clusters and cotranscribed; for example nearly 50 % of *X. tropicalis* microRNA genes are found in a microRNA cluster (Lee et al., 2002) (Tang and Maxwell, 2008). The oncogenic miR-17-92 cluster encodes a transcript containing six microRNAs and is conserved in vertebrates (Mendell, 2008). MicroRNA loci are generally found in introns, and this is true of *Xenopus* microRNAs (Kim et al., 2009; Tang and Maxwell, 2008)).

MicroRNAs are transcribed as long primary-miRNA transcripts (pri-miRNA) by RNA polymerase II in the nucleus (Figure 1.5). pri-miRNA is co-transcriptionally processed into ~ 70 nt hairpin precursors by a microprocessor complex consisting of Drosha and DGCR8, an essential dsRNA-binding protein cofactor. Drosha and DGCR8 recognise dsRNA:ssRNA junctions in pri-miRNA transcripts and cleave ssRNA from the dsRNA hairpin stem approximately 11 bp away from the dsRNA:ssRNA junction to form pre-miRNA (Han et al., 2006). Exportin5 then exports pre-miRNAs to the cytoplasm, where they interact with Dicer in the RNA-induced silencing complex (RISC). Dicer cleaves pre-miRNA at the opposite end of the hairpin to Drosha, removing the loop of the hairpin to form a miRNA:miRNA duplex (Macrae et al., 2006). The least thermodynamically stable strand of the miRNA duplex is then degraded, known as the miRNA* or 'passenger' strand. The mature, single-stranded miRNA then associates with Ago proteins.

Studies in *C. elegans* and *Drosophila* have shown that siRNA-like miRNA duplexes with perfect base-pairing are loaded onto Ago2, whereas miRNA duplexes with a central bulge from mismatched bases are loaded onto other members of the Ago family

(Ago1 in *Drosophila*) (Jannot et al., 2008) (Tomari et al., 2007). Ago2 contains endonuclease activity and can interact with other factors to cause exonucleolytic degradation of mRNA (Figure 1.6). However, this type of microRNA interaction with a target mRNA is rare in animals, and the human Ago proteins (Ago1-4) bind miRNAs without sorting by structure (Azuma-Mukai et al., 2008). GW182 proteins have recently emerged as critical regulators of microRNA interaction with target mRNAs. GW182 proteins contain an N-terminal Ago-binding domain with GW repeats and a C-terminal silencing domain and are required for miRNA-mediated mRNA degradation and inhibition of mRNA translation (Eulalio et al., 2008). The role of Ago proteins is to recruit GW182, as artificially tethering GW182 to the 3'UTR in the absence of Ago is sufficient to repress translation and cause mRNA degradation (Eulalio et al., 2008).

microRNA-mediated mRNA decay involves deadenylation of target mRNAs; for example miR-430 in zebrafish and its *X. laevis* homologue miR-427 mediate deadenylation of their respective maternal mRNA targets (Giraldez et al., 2006; Lund et al., 2009). Decapping follows removal of the poly(A) tail, which permits mRNA degradation by exonucleases in the 5' to 3' direction. GW182 interacts with poly(A)-binding protein (PABP) via the silencing domain to recruit the deadenylase complex CCR4/NOT1, thereby promoting mRNA degradation (carbon catabolite repression 4, negative on TATA-less)(Behm-Ansmant et al., 2006; Huntzinger et al., 2010). PABP promotes initiation of translation by interaction with the eIF4G scaffold protein (Figure 1.7), therefore GW182 interaction with PABP may disrupt the closed-loop conformation required for translation initiation. However, GW182 repression of translation does not require a poly(A) tail, so other interactions of GW182 may be important for translational repression (Eulalio et al., 2008).

Earlier models of miRNA-mediated translational repression focussed on Ago proteins binding the 5' methylguanosine cap to prevent eIF4E cap binding, which is required for initiation of translation (Figure 1.7) (Kiriakidou et al., 2007). This was supported by the observation that addition of recombinant eIF4E rescues miRNA-mediated repression of translation by let-7 in cell extracts (Mathonnet et al., 2007). However, later studies showed that functionally inactive Ago was still able to bind the cap (Eulalio et al., 2008). Impaired recruitment of the 60S ribosomal subunit was also suggested to account

for microRNA inhibition of translation, particularly as a complex containing the 60S subunit, Dicer and Ago was detected, however the mechanism is still unclear (Chendrimada et al., 2007).

Evidence from genome-wide proteomic studies and microarray analysis reveals that both mRNA stability and translation are affected in response to overexpression and knockdown of exogenous microRNAs (Baek et al., 2008; Selbach et al., 2008). Protein levels of hundreds of genes were modestly affected by manipulation of microRNA levels (generally less than 4-fold) (Selbach et al., 2008). While examples of translational repression without a decrease in mRNA levels were observed, in general, the strongest downregulation of protein levels following microRNA overexpression was accompanied by a decrease in mRNA levels (Baek et al., 2008; Selbach et al., 2008). Intriguingly, two measurements of protein and mRNA levels 24 hours apart showed that translation alone was inhibited first by microRNA overexpression and followed 24 hours later by a decrease in mRNA levels (and consequently translation), suggesting that mRNA degradation is a consequence of miRNA-mediated translational repression (Selbach et al., 2008).

Later studies measuring the effects of microRNAs on translation by sequencing mRNA associated with ribosomes (ribosome profiling) with parallel mRNA-seq transcriptome profiling showed that at least 84 % of decreased protein synthesis was due to lower mRNA levels (Guo et al., 2010). A study measuring mRNA association with polysomes (multiple ribosomes on one transcript) by sucrose gradient separation revealed that miR-124 target mRNA association with polysomes is decreased on miR-124 overexpression in human cells, suggesting that microRNAs repress translation by inhibiting the initiation step (Hendrickson et al., 2009). Again, modest effects on gene regulation were observed: mRNA levels of miR-124 targets decreased by 35 % and protein levels by 12 % (Hendrickson et al., 2009).

Genome-wide studies have also revealed that the major factor determining microRNA recognition of a target mRNA is the presence of an exact match to the microRNA seed sequence (nucleotides 2-7 of the miRNA) in the 3’ UTR of the mRNA (Baek et al., 2008; Hafner et al., 2010; Selbach et al., 2008). However, the contribution of

microRNA:mRNA base-pairing outside the seed region to microRNA target recognition has been documented (Bartel, 2009; Shin et al., 2010). The presence of mismatches in base-pairing between an mRNA containing a perfect seed match and nucleotides 9-11 of the miRNA correlates with repression of translation independent of mRNA level, suggesting that imperfect base pairing outside the seed region favours translational repression (Selbach et al., 2008). Ago proteins and microRNAs are known to bind coding regions as well as 3'UTRs; however, Ago binding sites in the coding region have minimal effects on mRNA stability in comparison to 3'UTR sites (Shin et al., 2010).

MicroRNA activation of TNF-α translation has been reported in quiescent cells and *Xenopus* oocytes and is dependent upon miR-369-3p-Ago binding to FXR1 (Fragile-X-related protein 1) (Mortensen et al., 2011; Vasudevan et al., 2007). The miRNA-Ago-FXR1 complex localises to polysomes for upregulation of translation in quiescent cells (Vasudevan et al., 2007). However, the TNF-α 3' UTR is a well-characterised site for repressive gene regulation mediated by RNA-binding proteins (see section 1.6.1), which could be affected by overexpression of miR-369-3p. In summary, microRNA activation of gene expression is in contrast to the genome-wide trend of miRNA repression of gene expression observed in proliferating cells.

Modulation of microRNA processing has emerged as a critical point of regulation to determine microRNA function. For example, Smads form a complex with the RNA helicase p68 to interact with Drosha to enhance processing of the TGF-β-responsive microRNA, miR-21 (Davis et al., 2008). A microRNA processing pathway that does not require Dicer has been identified in zebrafish and mice (Cheloufi et al., 2010a; Cifuentes et al., 2010). Instead, miR-451 is processed by the endonuclease activity of Ago2 which cleaves the miRNA* strand. The balance of Dicer or Ago2-mediated processing is determined by pre-miRNA structure. Dicer-dependent miRNAs can be converted into Ago-2 dependent miRNAs by shortening the pre-miRNA hairpin and removing internal bulges (Cifuentes et al., 2010). Whether this mechanism applies to a wider range of microRNAs has yet to be determined but it fits well with recent work that suggests limiting Ago protein restricts miRNA processing in *Xenopus* (Lund et al., 2011).

Some studies have suggested that in addition to their role of post-transcriptional regulation, microRNAs may also repress transcription by directing chromatin remodelling to the promoter, based on the observation that exogenous siRNAs complementary to the promoter can silence gene expression by this mechanism (see section 1.4.1). For example, the requirement for antisense transcription overlapping the promoter for siRNA-mediated chromatin remodelling of the INK4/ARF locus prompted the investigation of microRNAs complementary to promoter-overlapping transcripts (Gonzalez et al., 2008). Overexpression of miR-17-5p and miR-20a was shown to result in chromatin remodelling of promoters containing overlapping transcripts in mammalian cells, although how generally applicable this mechanism is has yet to be explored (Gonzalez et al., 2008). Another example of miRNA-directed chromatin remodelling is that of human miR-320, which is transcribed from the promoter of POLR3D (a cell cycle gene) in the antisense orientation and mediates AGO1 binding and acquisition of trimethylated H3K27 to silence POLR3D expression (Kim et al., 2008).

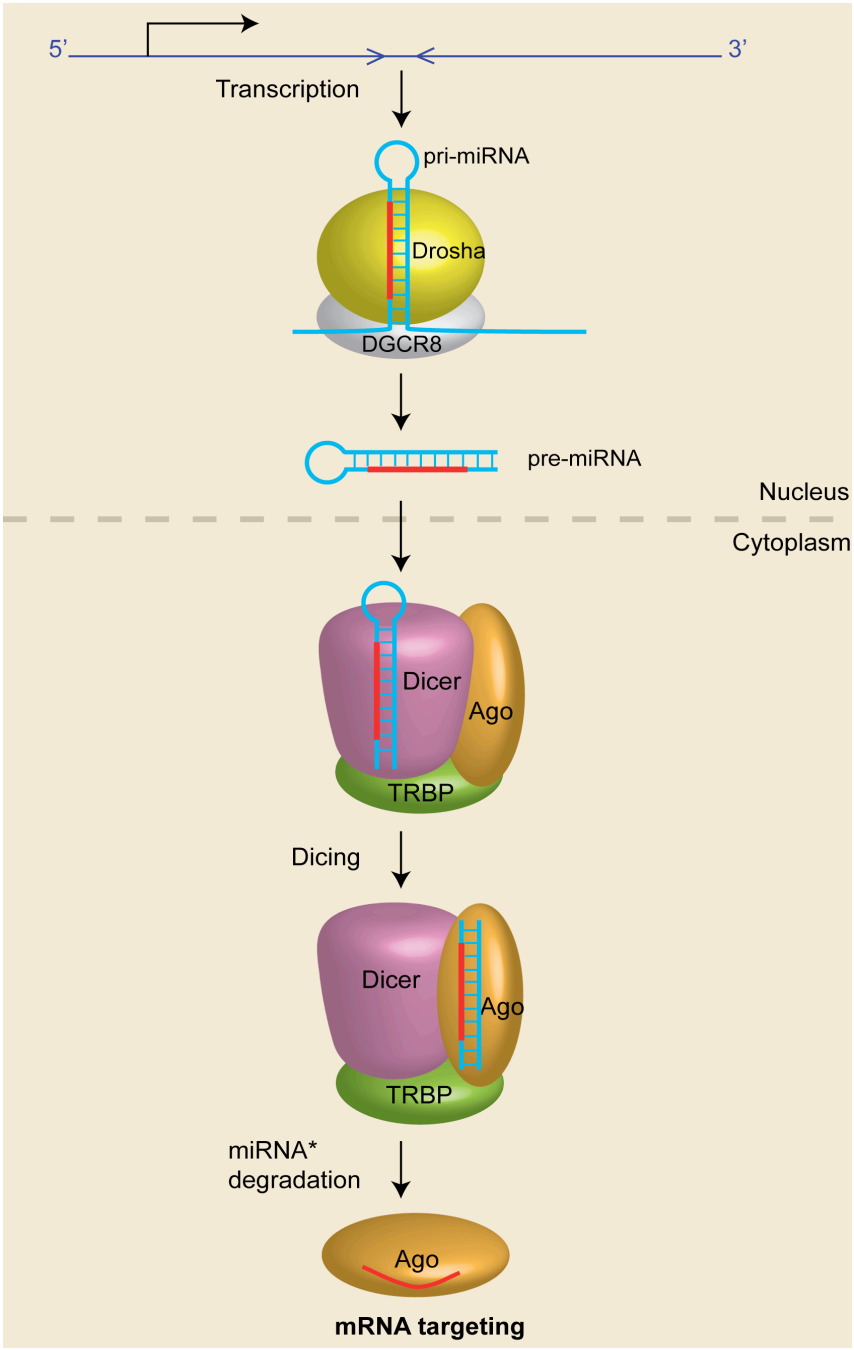


Figure 1.5 MicroRNA biogenesis pathway

MicroRNA genes are transcribed by RNA polymerase II in the nucleus and the primary-miRNA transcript is processed to a pre-miRNA hairpin by Drosha. The mature miRNA sequence is shown in red. Pre-miRNA is exported to the cytoplasm and is then 'diced' into a microRNA duplex by interaction with Dicer. The miRNA duplex is loaded onto Argonaute proteins and the miRNA* strand is degraded. Ago proteins (Ago1-4) mediate the interaction of the mature miRNA with its target mRNA. The dsRNA-binding proteins DGCR8 (DiGeorge syndrome critical region gene 8) and TRBP (transactivation-responsive RNA-binding protein) are cofactors for microRNA processing.

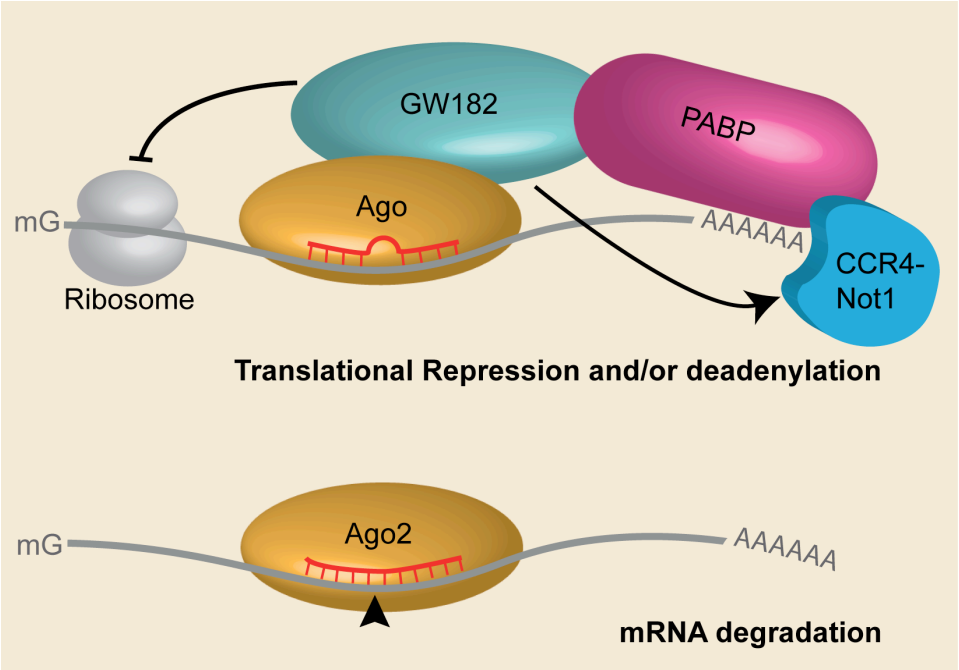


Figure 1.6 MicroRNA interactions with target mRNA

Mature microRNAs bind the 3'UTR of target mRNA to silence gene expression by two mechanisms: translational repression, or mRNA degradation. Mature miRNAs (red) bound to Ago proteins interact with GW182 proteins, which are required for mRNA degradation and repression of translation. GW182 promotes mRNA degradation by interacting with poly(A)-binding protein (PABP) to recruit deadenylases (CCR4-Not1 complex). GW182 proteins repress translation in a mechanism independent of the poly(A) tail or methylguanosine (mG) RNA cap binding (see Figure 1.7). Ago2 has slicer activity and can cause direct endonucleolytic cleavage of target mRNAs in some cases.

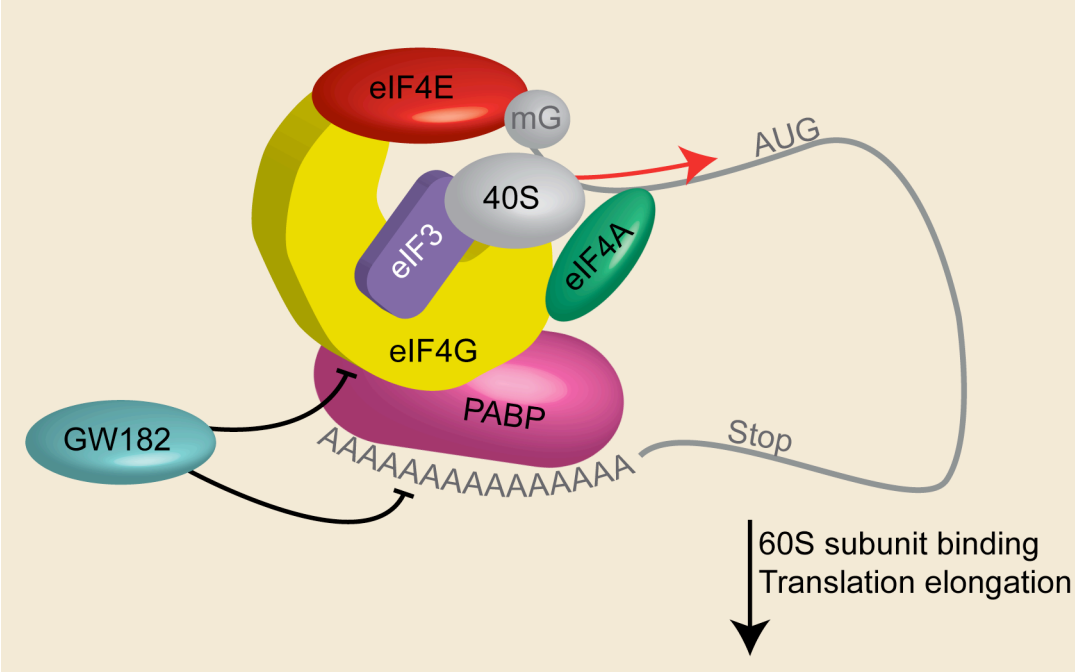


Figure 1.7 MicroRNA recruitment of GW182 to mRNA inhibits initiation of translation

Schematic of the translation initiation complex. The methylguanosine cap (mG) is recognised by the eIF4E translation initiation factor (eukaryotic initiation factor 4E), which interacts with eIF3 to recruit the 40S ribosomal subunit. eIF4G acts as a scaffolding protein, and interacts with PABP to form a circular loop of mRNA in the translation initiation complex. The 40S subunit scans the mRNA for the AUG start codon, indicated by the red arrow. GW182 proteins could repress translation via disruption of the PABP-eIF4G interaction, given that GW182 proteins interact with PABP. GW182 also promotes removal of the poly(A) tail to destabilise formation of the translation initiation complex.

1.4.3 piRNA biogenesis and mechanism of action

Piwi-interacting RNAs are 25-30 nt small RNAs that function via interaction with Piwi proteins in the Argonaute family and were initially discovered in *Drosophila* by small RNA sequencing (Aravin et al., 2003). Mutations in the *Drosophila piwi* gene (P-element induced wimpy testis) lead to male sterility and a lack of germ cells in the ovary (Lin and Spradling, 1997), pointing to a role of piRNAs in germline development. piRNAs are complementary to intergenic repeats (such as transposons) and mediate silencing of transposons (see section 1.5.3). piRNAs have been sequenced in mice, zebrafish and *X. tropicalis* (Aravin et al., 2006; Armisen et al., 2009; Houwing et al., 2007), and most recently in somatic tissues of the mouse and rhesus macaque (Yan et al., 2011).

In contrast to microRNAs, piRNAs do not require Dicer for their biogenesis (Vagin et al., 2006) and are modified at the 3' end by methylation of 2' hydroxyl groups catalysed by homologues of the *hen1* methyltransferase (Kamminga et al., 2010; Kirino and Mourelatos, 2007). There is evidence from zebrafish *hen1* mutants to suggest that methylation of piRNAs is critical for their stability (Kamminga et al., 2010). The components needed for piRNA biogenesis are largely uncharacterised, however the putative mechanism is described in Figure 1.8. The *Drosophila* Piwi family proteins are PIWI, AUB (Aubergine) and AGO3. piRNAs are transcribed from both strands of repetitive intergenic loci, are processed to 25-30 nt small RNAs by an unknown factor and subsequently associate with Piwi family proteins. The antisense piRNAs guide the Piwi proteins (which contain slicer activity) to sense transposon transcripts to mediate their cleavage and degradation, and initiate a 'ping-pong' cycle of antisense piRNA amplification (Brennecke et al., 2007; Gunawardane et al., 2007). Cleavage of the sense transposon transcript generates a sense piRNA, which then interacts with AGO3 to cleave antisense transposon transcripts and generate more antisense piRNAs (Figure 1.8). piRNA biogenesis varies between germline cells and somatic cells in *Drosophila*. Ping-pong amplification of piRNAs (see Figure 1.8) occurs in the germline to target a wide range of transposons transcripts for cleavage, whereas in somatic cells of the ovary, piRNA precursors are directly processed from the *flamenco* piRNA locus without an amplification step (Malone et al., 2009).

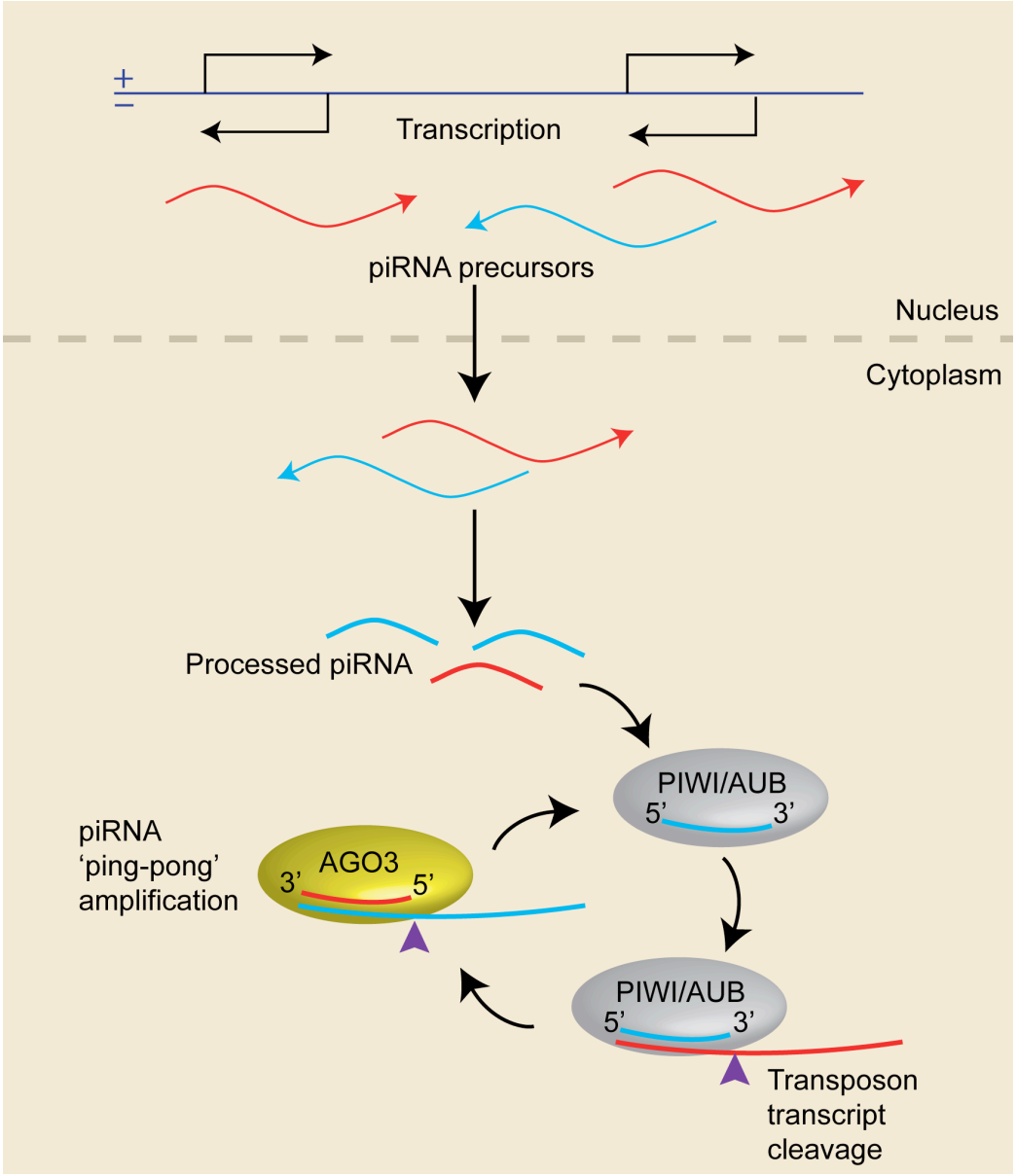


Figure 1.8 piRNA biogenesis requires processing and 'ping-pong' amplification
Schematic of piRNA synthesis in *Drosophila*. piRNA precursors are transcribed from repetitive elements on both strands of the genome and are processed into 25-30 nt lengths. piRNAs interact with Piwi family proteins PIWI or AUB (Aubergine) and recruit transposons with a complementary sequence. Endonuclease activity of Piwi family proteins is thought to cleave the transposon transcript (shown by a purple arrowhead), generating the 5' end of a new sense piRNA, which interacts with the Piwi protein AGO3 and in turn recruits an antisense transposon transcript. Subsequent cleavage of the antisense transposon transcript by AGO3 generates an antisense piRNA, which can bind PIWI/AUB to restart the cycle of piRNA amplification and transposon transcript cleavage.

1.5 Overview of small RNA function in early vertebrate development

1.5.1 siRNA function in early development

As discussed in section 1.4.1, exogenous and endogenous siRNAs can elicit RNAi knockdown of complementary transcripts in most organisms and direct heterochromatin formation to silence gene expression in plants, yeast and in certain circumstances, in mammalian cells. RNAi in plants and *C. elegans* is thought to be a viral defence mechanism: in response to viral DNA or RNA, complementary RNA is synthesised by an RNA-dependent RNA polymerase to form dsRNA duplexes which are then degraded (Ding and Voinnet, 2007; Wang et al., 2010a). However, the role of endogenous siRNAs in vertebrate animals with a complex immune system is not fully understood.

Sequencing of small RNAs associated with AGO1 and AGO2 in *Drosophila* ovary and somatic tissues revealed that endo-siRNAs mapped to sites of overlapping transcription, hairpin transcripts, transposons and piRNA clusters and required Dicer-2 for processing (Chung et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008). Mutation of Ago-2 or Dicer-2 also increased levels of transposon transcripts, suggesting that endo-siRNAs may play a similar role in somatic cells similar to that of piRNAs in transposon silencing in the germ line (Chung et al., 2008). Moreover, segmentation of Ago-2 and Dicer-2 mutants is defective when external temperature differences are applied to half the embryo, suggesting that endo-siRNAs could modulate gene expression in response to environmental conditions (Lucchetta et al., 2009).

Endo-siRNAs have recently been sequenced in *X. tropicalis* oocytes and somatic tissues, and in mouse ovaries and embryonic stem cells (Armisen et al., 2009; Babiarz et al., 2008; Watanabe et al., 2008). In these studies, mouse endo-siRNAs were shown to originate from overlapping transcription, hairpin transcripts and transposons, like *Drosophila* endo-siRNAs (Watanabe et al., 2008). An additional source of endo-siRNAs from pseudogenes was also identified (Watanabe et al., 2008). Downregulation of Dicer and Ago2 in mouse ovaries caused an increase in retrotransposon transcripts and in protein coding mRNAs, showing that mammalian endo-siRNAs can negatively regulate gene expression and protect against transposons (Watanabe et al., 2008). Since embryonic stem cells are models of early embryonic development, it is possible that the

role of mouse endo-siRNAs is conserved in the embryonic development of other vertebrate species.

1.5.2 MicroRNA function in early development

As discussed in section 1.4.2, microRNAs can negatively regulate gene expression by promoting mRNA degradation and/or inhibiting translation. Overexpression of a single microRNA can (modestly) downregulate the expression of hundreds of proteins (Selbach et al., 2008). An example of a well-characterised microRNA with multiple roles in early vertebrate development is miR-430 in zebrafish. Ectopic expression of miR-430 can rescue defects in brain morphogenesis, somite boundary formation and trunk morphology defects in embryos lacking Dicer (Giraldez et al., 2005). miR-430 also facilitates the onset of the zygotic transcriptional program at the mid-blastula transition in zebrafish by promoting the deadenylation and degradation of hundreds of maternal mRNAs (Giraldez et al., 2006). miR-430 processing requires Dicer, and so the onset of the maternal-zygotic transition is delayed in zebrafish mutants lacking Dicer (Giraldez et al., 2006). The *Xenopus* homologue of miR-430, miR-427 also regulates the stability of maternal mRNAs, although this has only been demonstrated for two targets, *cyclin A1* and *cyclin B2* (Lund et al., 2009).

Zebrafish miR-430 also contributes to regulation of genes specifically expressed in the germline. For example, in somatic cells, miR-430 destabilises *nanos1* mRNA, a gene required for germ cell development, whereas in germ cells, miR-430 activity is counteracted by RNA-binding proteins (see section 1.6), relieving miR-430-mediated repression of *nanos1*. Zebrafish miR-430 also regulates Nodal signalling by targeting the Nodal ligand *squint* and the inhibitor, *lefty* (Choi et al., 2007). Similarly, *Xenopus* miR-427 negatively regulates expression of Xnr5, Xnr6 and Lefty in *Xenopus* (Rosa et al., 2009). The Nodal pathway is also targeted at the receptor level by miR-15 and miR-16 in *Xenopus* (see section 1.1.2).

In addition to regulation of Nodal signalling, microRNAs have been implicated in regulation of multiple pathways in early development. *Xenopus* miR-449 and its human homologue target the Notch1 receptor and its ligand DLL1 (Delta-like 1) (Marcet et al., 2011). This interaction is critical for downregulation of Notch signalling in cilia

development on epithelial cells in humans and the epidermis in *Xenopus,* where miR-449 is first expressed in multi-ciliated progenitor cells at neurula stage (Marcet et al., 2011). In humans, transcription of miR-449 is under cell-cycle control and mature miR-449 inhibits cell proliferation (Bou Kheir et al., 2011; Yang et al., 2009). Loss of miR-449 has been documented in gastric tumours, pointing to a role for miR-449 as a tumour suppressor (Bou Kheir et al., 2011).

miR-214 in zebrafish controls muscle cell fate by modulation of Hedgehog signalling via direct targeting of *su(fu)* mRNA (Flynt et al., 2007). su(fu) negatively regulates Hh signalling by tethering the effector Gli (a transcription factor) to the cytoplasm, preventing transcriptional regulation of Hh target genes in the nucleus. miR-214 targeting of *su(fu)* amplifies the effect of high or low Hedgehog signalling on muscle cell fate (Flynt et al., 2007). Moreover, by targeting mRNA of the transcriptional repressor *hairy1*, *X. tropicalis* miR-9 can promote neurogenesis in the *X. tropicalis* embryo via FGF and Wnt-mediated proliferation of neural progenitors (Bonev et al., 2011).

The examples of microRNA control of embryonic signalling discussed point to a role of microRNAs in controlling cell differentiation. This is underscored by the tissue-specific expression of microRNAs observed in zebrafish and in *Xenopus tropicalis*, in particular in the nervous system (Walker and Harland, 2008; Wienholds et al., 2005). For example, *Xenopus* miR-124 is specifically expressed in the nervous system and restricts eye and forebrain neurogenesis via targeting the pro-neural transcription factor, NeuroD1, whilst also promoting cell proliferation (Liu et al., 2011). miR-9 and miR-124 are critical regulators of neural fate; for example, their ectopic expression can convert human fibroblasts to neurons (Yoo et al., 2011). miR-1 and miR-133 are specifically expressed in somites and muscle tissue in *X. tropicalis* and mouse embryos (Chen et al., 2006; Walker and Harland, 2008). miR-1 and miR-133 are cotranscribed in mammalian cells and regulate skeletal muscle development, illustrating the conservation of miRNA expression and function in vertebrates (Chen et al., 2006).

As mentioned above, in early *Xenopus* embryos and in human embryonic stem cells (a model of human embryonic development), miR-427 and its human homologue miR-

302 promote mesodermal cell fate via regulation of Nodal signalling (Rosa et al., 2009). Downregulation of miR-427/miR-302 leads to the loss of mesendodermal marker expression in both systems, and overexpression of miR-302 upregulates neuroectoderm markers *Pax6*, *Sox1* and pluripotency markers *Oct4* and *Nanog* (Rosa et al., 2009). This supports the suggestion that the function of microRNAs in early vertebrate development is to promote differentiation and repress pluripotency. MicroRNA downregulation (notably by coding region targeting) of *Nanog*, *Oct4*, and *Sox2* translation in mouse embryonic stem cells in response to retinoic acid-induced differentiation further supports this view (Tay et al., 2008). Moreover, microRNA expression profiles can distinguish the differentiation status of tumours and downregulation of global microRNA expression in tumours compared to normal tissues has been reported, suggesting that abrogation of microRNA-induced differentiation contributes to cancer (Lu et al., 2005).

The wider role of microRNAs in early vertebrate development has been assessed by knockdown or mutation of individual microRNAs and the microRNA-processing enzyme, Dicer. Whereas mouse Dicer knockouts are lethal, zebrafish *dicer* mutants display modest morphology and brain development defects (Bernstein et al., 2003; Giraldez et al., 2005). Similarly, zebrafish *ago2* mutants show delayed erythropoesis, whereas homozygous mouse *Ago2* mutants die as neonates due to anaemia, in both cases due to loss of the Dicer-independent microRNA miR-451 (Cheloufi et al., 2010a; Cifuentes et al., 2010). *Xenopus Dicer* morphants display late-development phenotypes. Eye development is inhibited and neural crest derivatives such as melanocytes and craniofacial cartilage are reduced (Gessert et al., 2010). This suggests that mammalian development is more sensitive to defects in mature microRNA production than zebrafish or *Xenopus* development or there are multiple *Dicer* genes in *Xenopus* and zebrafish. Mouse microRNA knockouts have revealed microRNA regulation of cardiac development (miR-1 and miR-208), immune cell development (miR-155 and miR-223) and miRNA requirement in early development (miR-17-92 cluster) (Smibert and Lai, 2008). Recently, mouse ES cells carrying deletions for 392 of the 476 mouse microRNA genes have been generated as a resource for miRNA knockout mouse generation, which should uncover the biological role of microRNAs in vertebrates (Prosser et al., 2011).

MicroRNAs have also emerged as critical regulators of epithelial to mesenchymal transition (EMT), a process crucial for morphogenesis and cell migration in development and metastatic spread of cancer, wherein cells undergo a change from epithelial to mesenchymal morphology (Chaffer and Weinberg, 2011). The zinc-finger transcription factors ZEB1and ZEB2 are critical activators of EMT and are transcriptional targets of TGF-β signalling (Comijn et al., 2001). ZEB1 and ZEB2 repress the expression of epithelial genes including *E-cadherin* by interaction with transcriptional corepressors. The miR-200 family directly targets ZEB1 and ZEB2 to promote epithelial differentiation, however, ZEB1 and ZEB2 also repress the transcription of the miR-200 family miRNAs to create a feedback loop (Bracken et al., 2008; Burk et al., 2008).

Overexpression of the miR-200 family members miR-200c, miR-203 and miR-183 in pancreas carcinoma cells leads to downregulation of stem-cell markers such as *Bmi1*, indicating that ZEB1/2 repression of the miR-200 family promotes acquisition of stem cell characteristics in cancer cells (Wellner et al., 2009). This agrees with previous work showing EMT-induced expression of stem cell markers and mammosphere formation (a trait of stem cells) in human mammary epithelial cells (Mani et al., 2008) and fits well with the idea that microRNAs repress pluripotency in cancer as well as in embryonic development.

A possible mechanism for global microRNA downregulation and EMT promotion in cancer is the downregulation of Dicer by miR-103/107 in breast cancer (Martello et al., 2010a). MicroRNA-mediated repression of Dicer is suggested to downregulate the processing of mature miR-200 family microRNAs, thereby relieving post-transcriptional repression of ZEB1/2 to promote EMT. Metastatic breast cancer cells were found to have high miR-103/107 levels (Martello et al., 2010a). MicroRNA control of metastasis in breast cancer by miR-10b under the transcriptional control of the EMT activating transcription factor Twist has also been documented (Ma et al., 2007).

In summary, microRNAs control many aspects of early vertebrate development, including the clearance of maternal mRNAs at the maternal-zygotic transition,

modulation of embryonic signalling pathways such as Nodal signalling, promoting cell differentiation and controlling cell fate, (particularly in germ layer formation and neural differentiation) and regulating epithelial to mesenchymal transition. Aberrant regulation or reactivation of embryonic signalling in many of these processes is observed in cancer.

1.5.3 piRNA function in early development

Mutations in Piwi proteins cause defective germ cell production and the activation of transposons in *Drosophila*, zebrafish and mice (Carmell et al., 2007; Girard et al., 2006; Houwing et al., 2007; Kalmykova et al., 2005; Lin and Spradling, 1997). The evidence suggests that the role of piRNAs in germ cell maintenance is directly related to their role in genome defence against transposon integration: piRNA biogenesis pathway mutations lead to the accumulation of double strand breaks and the activation of the ATR/Chk2 DNA damage checkpoint in *Drosophila* (Klattenhoff et al., 2007). In addition to mediating degradation of transposon transcripts, mouse piRNAs are thought to direct CpG methylation of DNA to transcriptionally silence LINE and LTR retrotransposons (Kuramochi-Miyagawa et al., 2008).

In addition to transposon silencing, *Drosophila* piRNAs are implicated in telomere protection: mutation of *aubergine* disrupts telomere binding to the telomere protection complex (Khurana et al., 2010). piRNAs can target repetitive transcripts that are not derived from transposons. For example, the protein coding *Stellate* gene contains multiple tandem repeats and is repressed by piRNA interaction with AUB2, which cleaves *Stellate* mRNA (Nishida et al., 2007). Moreover, piRNA targeting of maternal *nanos* mRNA for cleavage via recruitment of CCR4 deadenylase in a manner reminiscent of zebrafish miR-430 has been reported in *Drosophila* (Rouget et al., 2010). piRNAs from two transposons (*412* and *roo*) target 129 nt and 75 nt regions of the *nanos* 3' UTR respectively and form a complex with AUB, an RNA-binding protein (Smaug) and the deadenylase CCR4 to mediate poly(A) tail removal and subsequent degradation of *nanos* mRNA, which is critical for head development (Rouget et al., 2010).

The mechanism of piRNA interaction with chromatin is unclear. *Drosophila* Piwi family protein mutants (PIWI, AUB) and RNA-helicase *spindle-E* mutants (critical for piRNA synthesis) show loss of transposon silencing and heterochromatin including reduction in repressive epigenetic marks such as H3K9 methylation and decreased HP1 localisation (Klenov et al., 2007; Pal-Bhadra et al., 2004). HP1 binds methylated lysines and mediates silencing by recruitment of other factors such as histone deacetlyases and histone methyltransfereases in a repressive complex. These results suggest that piRNA biogenesis promotes heterochromatin formation at transposons. On the other hand, recent work has shown that heterochromatin directs piRNA synthesis. Methylation of H3K9 by the histone methyl-transferase dSETDB1 is essential for transcription of piRNA clusters and in its absence transposons are activated, leading to loss of germ cells (Rangan et al., 2011). Moreover, HP1 recruitment is independent of AGO2 and PIWI expression, suggesting that heterochromatin formation is independent of siRNA or piRNA-guided recruitment of these proteins to DNA (Moshkovich and Lei, 2010).

piRNAs were recently sequenced in the mouse pancreas and in tissues of the macaque reproductive system, suggesting that piRNAs may play a role outside of genome defence in the germline. Macaque piRNA expression domains overlapped with Piwi family mRNAs and expression of the mouse PIWI homologue in the pancreas was confirmed (Yan et al., 2011). The role of piRNAs outside of maintaining germline integrity could include silencing transcription from non-transposon repeats or maintenance of heterochromatin, although the best understood function of human and mouse piRNAs is in germline development (Girard et al., 2006). Similarly, *Xenopus* Piwi family proteins are expressed in testes and oocytes, and piRNAs are the most abundant class of small RNAs found in oocytes and mature eggs (Lau et al., 2009). Both *Xenopus* Piwi family proteins XIWI and XILI are found in oocytes but are not expressed in adult *Xenopus* tissues (Armisen et al., 2009), suggesting that piRNA biogenesis is not active in adult *X. tropicalis* animals. In summary, piRNAs are critical for maintaining genome integrity in the germline of animals, but may have other functions, particularly in somatic tissues, which have yet to be explored.

In summary, small non-coding RNAs have emerged in the last decade as critical regulators of gene expression in early development. Recent research has also shown that

interplay between small RNAs and RNA-binding proteins in the 3' UTR of their target genes is an important mechanism for modulation of gene expression in embryogenesis.

1.6 Role of RNA-binding proteins in early vertebrate development

1.6.1 Regulation of RNA stability

As discussed in section 1.4.2, mRNA degradation can be mediated by deadenylase-mediated removal of the poly(A) tail and removal of the 5' methyl-guanosine cap. GW182 interacts with poly(A)-binding protein (PABP) to recruit a deadenylase complex (see Figure 1.6) via the silencing domain which is similar to the PABP-interaction motif 2 (PAM2) of the PABP-interacting protein PAIP2 (Huntzinger et al., 2010). The PAIP proteins regulate PABP activity: PAIP1 promotes translation, whereas PAIP2 inhibits translation by displacing PABP from the poly(A) tail. The PAM2 motif is present in various RNA-binding proteins that regulate mRNA stability or translation such as eukaryotic peptide chain release factor 3A (ERF3A) (Albrecht and Lengauer, 2004).

Deadenylation of the poly(A) tail has been extensively studied in *Xenopus*, where the RNA-binding protein EDEN (embryonic deadenylation element) binds the 3' UTR of target oocyte mRNAs to clear them from fertilized embryos. The human homologue of EDEN is CUG-BP, which recruits the deadenylase PARN (poly(A)-specific ribonuclease) to initiate mRNA decay (Moraes et al., 2006). CUG-BP was initially identified for its propensity to bind CUG repeats but it also recognises AU-rich elements (AREs) in the 3'UTR. CUG-BP belongs to the *elav* family of RNA-binding proteins that are conserved in vertebrates and regulate mRNA stability and translation, named after the *Drosophila* mutant *embryonic lethal abnormal visual system* (Samson, 2008).

Whilst many *elav* proteins are neuronal, the human *elav* protein HuR is expressed ubiquitously, is nuclear and cytoplasmic and promotes stability of many mRNAs including *MyoD*, *p21*, *VEGF* and *TNF-*α amongst others (Barreau et al., 2005). HuR contains three RNA-binding domains and recognises a 17-20 bp U-rich motif, with

UAUUUAU as a core high-affinity motif (Lebedeva et al., 2011; Lopez de Silanes et al., 2004). Genome-wide studies of HuR binding revealed that as well as 3' UTR recognition, HuR binds in introns of pre-mRNA, indicating a possible role in pre-mRNA splicing consistent with previous results showing HuR association with the spliceosome (Chen et al., 2007; Lebedeva et al., 2011; Mukherjee et al., 2011). The number of HuR binding sites directly correlated with mRNA stability, and mRNAs with both intronic and 3' UTR HuR binding were the most stable (Mukherjee et al., 2011). HuR knockdown upregulates processing of miR-7, an intronic miRNA residing in a gene containing intronic and exonic Hur binding sites (Lebedeva et al., 2011). In this study, many HuR and Ago binding sites were found to be within 10 nt of each other and microRNA seeds were generally proximate but not overlapping with Hur binding sites, suggesting that HuR competes with microRNAs for 3' UTR binding (Lebedeva et al., 2011; Mukherjee et al., 2011).

Other ARE-binding proteins have a well-characterised role in promoting mRNA degradation. For example, tristetrapolin (TTP) targets the mRNA of *HIF-1α*, *TNF-α* and proteins involved in breast cancer invasion and metastasis including urokinase plasminogen activator (*uPA*) and matrix metalloproteinase-1 (*MMP-1*) for degradation (Al-Souhibani et al., 2010). TTP itself is downregulated whilst its targets are upregulated in the breast cancer cell line MDA-MB-231 compared to control breast cell lines, suggesting that TTP may play a role in breast cancer (Al-Souhibani et al., 2010).

RNA-binding proteins can also modulate clearance of maternal mRNAs by miR-430 in zebrafish. DAZL (deleted in azoospermia-like) is specifically expressed in primordial germ cells in zebrafish (which give rise to gametes) and promotes polyadenylation of target mRNAs to oppose the deadenylation activity of miR-430 (Takeda et al., 2009). In this way, DAZL promotes translation of the germ cell specific proteins nanos1 and tdrd7 (Takeda et al., 2009). Moreover, in *Xenopus*, DAZL interacts with PABP to promote translation initiation (Collier et al., 2005). DAZL is also essential for germ cell formation and differentiation of primordial germ cells into male or female gametes in the mouse (Gill et al., 2011).

DND1 (dead end 1) is also expressed in primordial germ cells in zebrafish and humans and counteracts miR-430 activity in fish by preventing recognition of miR-430 binding sites in the 3' UTR by the miR-430:RISC complex (Kedde et al., 2007). DND1 promotes RNA stability by recognition of U-rich regions close to miRNA binding sites via its RNA binding domain. Zebrafish DND1 targets include the germ cell specific genes *nanos* and *tdrd7*, whilst human target mRNAs include tumour suppressors *LATS2* and *p27*, which are negatively regulated by miR-372 and miR-221 respectively (Kedde et al., 2007). Essentially, DND1 binding prevents microRNA-RISC complexes binding microRNA target sequences in the 3' UTR to alleviate microRNA-mediated gene repression.

1.6.2 Translational control

Several RNA-binding proteins regulate translation efficiency without affecting mRNA levels. As described in Figure 1.7, translation initiation involves recognition of the methyl-guanosine cap by the translation initiation factor eIF4E to form a closed loop complex with mRNA via eIF4E interaction with eIF4G and PABP bound to the poly(A) tail. RNA-binding proteins containing an eIF4E-binding motif can compete with eIF4G for eIF4E binding and inhibit initiation of translation by preventing formation of the closed loop complex. For example, *Xenopus* cyclin B1 is translationally repressed in oocytes via interaction of Maskin, which contains an eIF4E-binding motif, with CPEB (cytoplasmic polyadenylation element binding protein) in the 3' UTR (de Moor et al., 2005). Translation of Maskin itself is activated by poly(A) tail elongation, which promotes interaction of PABP with eIF4G (Meijer et al., 2007). Similarly, *Drosophila oskar* and *nanos* mRNAs are translationally repressed by Bruno (a homologue of *Xenopus* EDEN and *elav* family member), which recruits the eIF4E-binding protein Cup to inhibit translation (Richter and Sonenberg, 2005). In contrast, the affinity of mammalian eIF4E binding proteins (4E-BP) for eIF4E is decreased by mTOR-mediated phosphorylation to link control of protein synthesis to cellular nutritional status and growth (Richter and Sonenberg, 2005). The *elav* family proteins ElrA and ElrB also mediate translational repression of Vg1 mRNA in the *X. laevis* oocyte by binding to AREs in the Vg1 3' UTR, however the mechanistic details are unclear.

Pumilio RNA-binding proteins also inhibit translation and are conserved from yeast to humans. Pumilio was first identified as a translational repressor of maternal *hunchback* mRNA in the posterior of the *Drosophila* embryo, but is now known to associate with many mRNAs via the 3' UTR motif UGUANAUA in *Drosophila* and humans (Gerber et al., 2006; Morris et al., 2008). Pumilio is able to repress cap-independent translation in *Drosophila,* but the mechanism of Pumilio translational inhibition remains unclear, although it is known to be dependent on residues in the RNA-binding domain (Wharton et al., 1998). Knockdown of a human Pumilio protein, Pum1 resulted in increased half-life for Pum1-associated mRNAs, in agreement with observations of Pumilio proteins in other species that Pumilio regulates RNA stability (Morris et al., 2008). Yeast Pumilio proteins promote mRNA deadenylation and degradation by interacting with components of the CCR4-Not deadenylase complex (Goldstrohm et al., 2006).

Recent work has indicated that Pumilio proteins can interact with microRNA regulation. Human Pum1 and Pum2 target mRNAs are enriched in microRNA binding sites, and evidence for Pumilio binding altering mRNA structure to enhance microRNA binding has been demonstrated for miR-221 and miR-222 binding of the p27 3'UTR in mammalian systems (Galgano et al., 2008; Kedde et al., 2010). Interactions between microRNAs and RNA binding proteins such as Hur, DND1 and Pumilio in 3' UTR-mediated gene regulation are emerging as an important mechanism for combinatorial control of gene expression.

In summary, early vertebrate development is orchestrated by the activation of transcriptional programs in response to extracellular ligands, modulation of signal transduction and control of gene expression by regulators including coreceptors such as XCR1, small RNAs including microRNAs, endo-siRNAs and piRNAs and RNA-binding proteins.

1.7 Aims of the thesis

- To explore the role of small RNAs as regulators in early vertebrate development.
- To investigate the mechanism and functional relevance of spatial regulation of *Xenopus* Cripto-1, a coreceptor for the Nodal signalling pathway.
- To characterise the small RNA profile of early vertebrate development by genome-wide sequencing of the small RNAome of the early *X. tropicalis* embryo.

Chapter 2. Materials and Methods

Laboratory chemicals were purchased from Sigma-Aldrich, Roche or Fisher Scientific unless otherwise stated. Standard laboratory solutions including 10 x TAE, 10 x TBE, 5 M sodium chloride, 0.5 M EDTA (pH 8), 1 M Tris-HCl (pH 7.8 or 8), 10 x PBS, 1 x PBS and autoclaved water were prepared by Cancer Research UK Cell Services.

2.1 Molecular Biology

A list of general molecular biology solutions and buffers is detailed in Section 2.1.3.

2.1.1 Preparation of plasmid DNA

2.1.1.1 Transformation of TG1 super-competent bacteria

Plasmid DNA from the Developmental Signalling laboratory plasmid database or gifts from other laboratories was retransformed into competent TG1 bacteria. 50 µl aliquots of TG1s were thawed on ice and incubated with 100 ng of plasmid DNA for 10 minutes, then subjected to a mild heat shock by incubation at 37 °C for 1 minute, followed by incubation on ice for 2 minutes. 200 µl of LB broth was added and shaking incubation at 37 °C for 20 minutes was carried out to allow for recovery of the TG1s and activation of the ampicillin resistance gene. Transformed bacteria were then plated onto pre-warmed LB-agar plates containing ampicillin at a concentration of 100 µg/ml to select for transformed bacteria. After overnight incubation at 37 °C, a single colony was picked using a sterile pipette tip and added to a flask containing 200 ml of LB broth supplemented with 100 µg/ml ampicillin and incubated shaking at 37 °C overnight to grow large-scale transformed bacterial culture from which to extract plasmid DNA.

2.1.1.2 Maxi-prep plasmid DNA purification

Plasmid DNA was extracted from large-scale bacterial cultures using the GenElute kit (Sigma) according to the manufacturer's instructions. Briefly, bacterial cultures were centrifuged at 4000 rpm for 20 minutes to obtain a bacterial pellet, which was resuspended and bacteria were then lysed using sodium hydroxide and SDS lysis buffer to release plasmid DNA. Subsequent column purification was used to remove

contaminating cellular debris from the DNA that was bound to the silica column membrane, which was then washed with ethanol, eluted and precipitated using isopropanol. After ethanol washes, the DNA pellet was air-dried and resuspended in TE (see section 2.1.3).

2.1.1.3 Estimation of nucleic acid concentration

The concentration of purified plasmid DNA was estimated by quantification on a Nanodrop spectrophotometer. This calculates the DNA concentration, **c**, in ng/μl based on the absorbance, **A**, of 260 nm UV light passed through the nucleic acid sample using the Beer-Lambert law. At 260 nm, double-stranded DNA has an average extinction coefficient (**ε**) of 0.02 $(ng/\mu l)^{-1}cm^{-1}$, and the path length, **b**, in the Nanodrop is 10 mm. The Beer-Lambert law can be rearranged to calculate DNA concentration as follows:

$$c = \frac{A}{\varepsilon b}$$

For example, an absorbance of 1 corresponds to a DNA concentration of 50 ng/μl. Absorbance ratios at different wavelengths can indicate the purity of the DNA sample. The A260/A280 ratio is 1.8 for pure DNA. For purified RNA, a sample with an absorbance of 1 at 260 nm corresponds to an RNA concentration of 40 ng/μl. The A260/A280 ratio for pure RNA is 2.0.

2.1.2 Cloning

2.1.2.1 PCR amplification and product purification

The sequence of interest was amplified from template DNA by the polymerase chain reaction (PCR). PCR forward and reverse primers contained overhanging sites for restriction enzymes to allow cloning of the PCR product into a digested vector (see section 2.1.6 for a list of primers used in cloning). A typical PCR reaction and program are described below. Homemade Taq polyermase was prepared by Mike Howell. The PCR reaction was then subjected to agarose gel electrophoresis (1 % or 2 % agarose depending on expected product sizes) and visualised by ethidium bromide staining to confirm the sizes of the PCR products by comparison to standard DNA markers (NEB). The band was then excised from the gel and DNA was extracted from the gel slice using

the QIAquick Gel Extraction kit (Qiagen) using the microcentrifuge protocol according to the manufacturer's instructions. Briefly, this involved melting and solubilising the gel slice, applying the solution to a column that bound the DNA, ethanol washing and DNA elution from the column by adding 30 μl water heated to 50 °C and centrifugation.

PCR reaction (25 μl)

10 x PCR buffer 2.5 μl
10 μM forward primer 0.5 μl
10 μM reverse primer 0.5 μl
DNA (100 ng/μl) 1 μl
10 mM dNTPs 0.5 μl
water 19 μl
2.5 units Taq (homemade) 1 μl

94 °C 2 minutes
94 °C 30 seconds
50 °C 30 seconds
72 °C 30 seconds (or 1 minute/kB)
} x 30 cycles
72 °C 5 minutes
4 °C ∞

2.1.2.2 Restriction digests

The gel purified PCR product was then digested with the intended restriction enzymes (NEB), gel purified again and the DNA concentration was determined by Nanodrop quantification. The vector was prepared by restriction digest of 3 μg of DNA and 5' phosphates were removed by incubation with alkaline phosphatase (Roche) prior to gel purification and DNA quantification. Where the selected restriction enzyme digest would be inhibited or blocked by Dam methylation (e.g. XbaI), the DNA was retransformed into the Dam negative bacterial strain JM110.

Insert restriction digest

30 µl gel purified DNA

4 µl 10 x BSA

4 µl NEB reaction buffer

1 µl restriction enzyme A (20,000 units/ml)

1 µl restriction enzyme B (20,000 units/ml)

Incubate for 2 hours at 37 °C

Vector restriction digest

3 µg of vector DNA

2 µl 10 x BSA

2 µl NEB reaction buffer

1 µl restriction enzyme A (20,000 units/ml)

1 µl restriction enzyme B (20,000 units/ml)

Up to 20 µl with water

Incubation for 2 hours at 37 °C, followed by incubation with 1 µl (20 units/µl) of alkaline phosphatase (Roche) for 5 minutes at room temperature.

2.1.2.3 Ligations and transformation

The prepared insert and vector DNA were ligated using T4 DNA ligase (NEB). A vector-only control was included to assess the extent of vector re-ligation. Ligation reactions were incubated for 1 hour at room temperature and then transformed into TG1 bacteria, plated onto LB-agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight to allow colonies to grow.

2.1.2.4 Screening colonies by restriction digest and confirmation of successful cloning by sequencing

Colonies were picked using a sterile pipette tip and added to 3 ml of LB medium supplemented with the relevant antibiotic and incubated overnight at 37 °C in a shaking

incubator. Plasmid DNA was extracted from the small-scale bacterial cultures using the QIAprep Spin Mini-prep kit (Qiagen) according to the manufacturer's instructions. Briefly, bacteria were lysed using an alkaline lysis buffer and lysates were neutralised and cleared of cellular debris by centrifugation. The solution containing a white flocculate was then applied to a silica column that binds DNA. The column was washed with ethanol and DNA was eluted by the addition of 50 μl water and centrifugation. The DNA samples were then screened for presence of the insert by restriction digest. 8 μl of mini-prep DNA (concentration approximately 40 ng/μl) was digested with the selected restriction enzymes and the reaction was subjected to electrophoresis on a 1 % agarose gel. At least one positive sample was retransformed and a single colony was picked and grown up in a large-scale bacterial culture. DNA was extracted as in Section 2.1.1.2, and sequenced as described below. The sequencing reaction was cleaned up to remove unincorporated dye-terminators using the Dyex 2.0 Spin Kit (Qiagen) and the DNA was dried using a vacuum centrifuge. Samples were then submitted to the Cancer Research UK Equipment Park for capillary sequencing.

Sequencing reaction

1 μl DNA (200 ng)
1 μl primer (10 μM)
8 μl (Applied Biosystems) Big Dye Terminator
10 μl water

Sequencing program

96 °C 10 seconds
50 °C 5 seconds
60 °C 4 minutes
60 °C 1 minute
} x 30 cycles
10 °C ∞

2.1.3 General molecular biology solutions and buffers

500 x Ampicillin

2.5 g Ampicillin (Sigma)

50 ml water

Ampicillin stocks were aliqoted and stored at -20°C and used at a final concentration of 100 μg/ml

TE

10 mM Tris-HCl pH 7.5

1 mM EDTA

Autoclaved and stored at room temperature.

10 x PCR buffer

500 mM KCl

100 mM Tris-HCl pH 8.3

15 mM $MgCl_2$

Aliquoted and stored at -20 °C.

10 mM dNTPs

100 mM dATP 10 μl

100 mM dCTP 10 μl

100 mM dGTP 10 μl

100 mM dTTP 10 μl

autoclaved water 60 μl

dNTPs were from Roche and stored at -20 °C.

10 x TAE

48.4 g of Tris base

11.4 ml glacial acetic acid

20 ml 0.5 M EDTA (pH 8.0)

Up to 1 L with water

This solution was prepared and autoclaved by Cancer Research UK Cell Services and stored at room temperature.

6 x DNA loading dye

0.25 % (w/v) bromophenol blue (Sigma)

0.25 % (w/v) xylene cyanol (Sigma)

30 % (w/v) glycerol (Fisher)

1 x TAE

1% agarose gel mixture

1 g agarose (Invitrogen)

100 ml 1 x TAE

5 µl of 20,000 x ethidium bromide was added to the molten agarose before pouring the gel

20, 000 x Ethidium bromide

100 mg ethidium bromide (Sigma)

10 ml autoclaved water

10 x BSA

100 µl 100 x BSA (10 mg/ml, NEB)

900 µl water

Aliquoted and stored at - 20 °C.

2.1.4 PCR-based site directed mutagenesis

PCR primers used for site-directed mutagenesis were designed to contain the intended mutations flanked by 15 nucleotides of wild type sequence. For each mutant, the first round of PCR generates two overlapping products containing the desired mutation (Figure 2.1). These products were gel purified and 1 μl of each PCR product was used as the template for a second round of PCR using a pair of outer primers (containing restriction sites) to amplify the template DNA containing the desired mutations. A list of oligos used in site directed mutagenesis is detailed in section 2.1.6. The PCR product was then digested, gel purified and ligated into the digested vector as described in Sections 2.1.2.1-3.

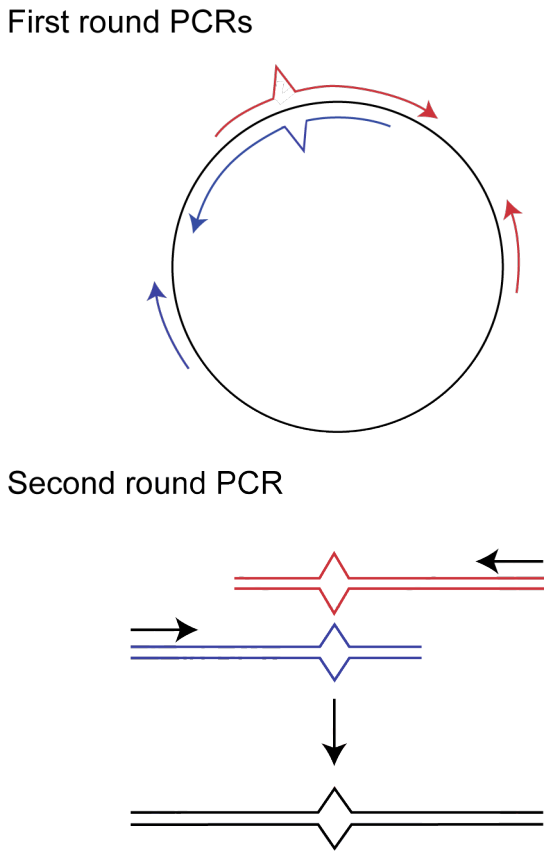


Figure 2.1 Site directed mutagenesis overview

2.1.5 List of plasmids

Plasmid Name	Description	Plasmid made by/ gift from	Reference
pSP72-XEF1α	*Xenopus* EF1α for RNAse protection probe (120 nt). Linearise with HinfI and transcribe with SP6 polymerase.	J. Smith	(Krieg et al., 1989)
pGEM-FRL1α	*Xenopus* XCR1α coding region (573 nt) for Northern blot probe. Linearise with SpeI and transcribe with T7 polymerase.	K. Dorey	(Dorey and Hill, 2006)
XFGF-R	*Xenopus* FGF-receptor for RNAse protection probe (146 nt). Linearise with AseI and transcribe with T7 polymerase.	C. Hill	(Howell et al., 2001)
pGEM3Z-XCR1α/β probe	1-250 nt of the *Xenopus* XCR1α coding region for an RNAse protection probe to detect both XCR1α (250 nt protected fragment) and XCR1β (200 nt protected fragment) mRNA. Linearise with XbaI and transcribe withT7 polymerase.	J. Harding	Unpublished
pGEM3Z-GFP probe	190 nt GFP coding region for an RNAse protection probe. Linearise with XbaI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-FRL1α I-tag	Coding region of XCR1α with an internal Flag-tag introduced after Leu37 for making sense mRNA for overexpression. Linearise with XbaI and transcribe with T7 polymerase.	K. Dorey	(Dorey and Hill, 2006)
pBSK(-)XHex	*Xenopus* Hex for *in situ* probe. Linearise with NotI and transcribe with T7 polymerase.	A. Zorn	(Newman et al., 1997)
pSP73-Xbra	*Xenopus* brachyury full length cDNA (2210 bp) for *in situ* probe. Linearise with BglII and transcribe with T7 polymerase.	J. Smith	(Smith et al., 1991)

pBSKS-XGsc	*Xenopus* goosecoid cDNA for *in situ* probe (440 nt). Linearise with XbaI, and transcribe with T3 polymerase.	J. Smith	(Cho et al., 1991)
pGEM4Z-Forkhead	*Xenopus* forkhead fragment for *in situ* probe. Linearise with EcoRI and transcribe with T7 polymerase.	M. Howell	(Howell and Hill, 1997)
pBSKII-Xantivin	*Xenopus* antivin (lefty) full length cDNA (1.9 kB) for *in situ* probe. Linearise with BamHI and transcribe with T7 polymerase.	M. Wright	(Cheng et al., 2000)
pBSK(-) Xsox17α	*Xenopus* sox17α full length cDNA (1.2 kB) for *in situ* probe. Linearise with KpnI and transcribe with T3 polymerase.	A. Zorn	(Sinner et al., 2004)
ALK4 *	Constitutively-active mouse ALK4 for sense mRNA synthesis. Linearise with HindIII and transcribe with T7 polymerase.	J. Smith	(Armes and Smith, 1997)
pFTX4K-Xnr5	*Xenopus* full length Xnr5 for sense mRNA synthesis. Linearise with XbaI and transcribe with T7 polymerase.	K. Dorey	(Dorey and Hill, 2006)
pFTX4K-Xnr6	*Xenopus* full length Xnr6 for sense mRNA synthesis. Linearise with XbaI and transcribe with T7 polymerase.	K. Dorey	(Dorey and Hill, 2006)
pCS2+GFP	GFP coding region for sense mRNA synthesis. Linearise with NotI and transcribe with SP6.	P. Blader	
pCS2+GFP-Dicer-MO	*X. laevis* Dicer morpholino binding site sequence directly upstream of the GFP coding region to check the morpholino works. Linearise with NotI and transcribe with SP6 polymerase for sense RNA.	S. Gessert	(Gessert et al., 2010)
pCS2+N-mCherry-H2B	Human histone H2B (full length) downstream of mCherry for sense mRNA. Linearise with NotI and transcribe with SP6 polymerase.	S. Woolner	Unpublished
pFTX4K-eGFPC1	GFP coding region with a stop-codon oligo at the 3' end of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished

pFTX4K-eGFPC1-XCR1α-3'UTR	GFP-3'UTR reporter. Full length XCR1α-3'UTR (1-958 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1β-3'UTR	GFP-3'UTR reporter. Full length XCR1β-3'UTR (1-764 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1β 300-764 3'UTR	GFP-3'UTR reporter. Partial XCR1β-3'UTR (1-764 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-Xtcr1 3'UTR	GFP-3'UTR reporter. Full length *X. tropicalis* Cripto-1 3'UTR (1-783 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-hCR1-3'UTR	GFP-3'UTR reporter. Full length human Cripto-1 3'UTR (1-1212 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-XCR1α 1-150 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (1-150 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 1-200 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (1-200 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 1-350 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (1-350 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished

pFTX4K-eGFPC1-XCR1α 392-958 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (392-958 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 286-637 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (286-637 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 276-402 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-402 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 276-461 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-461 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 276-579 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-579 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 450-520 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (450-520 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished
pFTX4K-eGFPC1-XCR1α 392-520 3'UTR	GFP-3'UTR reporter. Partial XCR1α-3'UTR (392-520 nt) cloned downstream of GFP. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	J. Harding	Unpublished

pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-107	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP with 5'TGCTGC to TCCACG mutations introduced beginning at 285 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-499	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP with 5'AGTCTTA to TCAGTAA mutations introduced beginning at nt 310 in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-gga-miR-1810	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP with 5'TATTAG to ATATAC mutations introduced beginning at 360 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-Pub1	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP with 5' TTTTTTTTTAATTTTTTT to TATGAACTTAATCAATAC mutations introduced beginning at 404 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-140	GFP-3'UTR reporter. Partial XCR1α-3'UTR (276-520 nt) cloned downstream of GFP with 5' TGTGGT to AGTCCA mutations introduced beginning at 489 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished

pFTX4K-eGFPC1-mut-miR-107-full length XCR1α 3'UTR	GFP-3'UTR reporter. Full length XCR1α-3'UTR cloned downstream of GFP with 5'TGCTGC to TCCACG mutations introduced beginning at 285 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished
pFTX4K-eGFPC1-mut-miR-1810-full length XCR1α 3'UTR	GFP-3'UTR reporter. Full length XCR1α-3'UTR cloned downstream of GFP with 5'TATTAG to ATATAC mutations introduced beginning at 360 nt in the 3'UTR. For sense mRNA, linearise with SpeI and transcribe with T7 polymerase.	R. Randall	Unpublished

2.1.6 List of oligos used in cloning

Restriction sites and sequence that was removed by restriction digest is indicated in lower case. Red indicates the site of a mutation that was introduced into the plasmid by site directed mutagenesis.

Plasmid Name	Insert Primers
pGEM3Z-XCR1α/β probe	F: 5'gaatctagaGGCACGAGGCGACCGTTACA R: 5'caaggatccTCCATGGGTGGCATTAATTG
pGEM3Z-GFP probe	F: 5' gaatctagaGCTAGCGCTACCGGTCGCCA R: 5'caaggatccCACGGGCAGCTTGCCGGTGG
pFTX4K-eGFPC1	Stop oligo cloned into parental vector by BglII/KpnI digest. 5' gatctCTAGATAACTGATggtac 5' cATCAGTTATCTAGa
pFTX4K-eGFPC1-XCR1α-3'UTR	F: 5' catggatccCAGAGTAACTTGAGTCTG R: 5' gcgcatcgatAGGTACAGTATTTAATGTCT
pFTX4K-eGFPC1-XCR1β-3'UTR	F: 5' catccatccCTGAGTAACTTGAGTCTG R: 5' gcgcatcgatACTTGTTCAGTTTATATTT
pFTX4K-eGFPC1-XCR1β 300-764 3'UTR	F: 5' catggatccTAATCAGACACGTTGTACC R: 5'ggtactagtACTTGTTCAGTTTATATTT

pFTX4K-eGFPC1-Xtcr1 3'UTR	F: 5' gcgcggatccTGAGTAACTAGAGTCTGCTAG R: 5' cgcgatcgatTACGTATTTAATGTTTCATTGC
pFTX4K-eGFPC1-hCR1-3'UTR	F: 5' catggatccTCGACATTGACCTATTTC R: 5' cgtactagtTTAGGTCAATATAGTTTCC
pFTX4K-eGFPC1-XCR1α 1-150 3'UTR	F: 5'catggatccCAGAGTAACTTGAGTCTG R: 5' gcgcatcgatATAGTTTCTGCATATTTTCA
pFTX4K-eGFPC1-XCR1α 1-200 3'UTR	F: 5'catggatccCAGAGTAACTTGAGTCTG R: 5' gcggatcgatCCTATAAAAAAAGTTATTTTATGAAAC
pFTX4K-eGFPC1-XCR1α 1-350 3'UTR	F: 5'catggatccCAGAGTAACTTGAGTCTG R: 5'gcgcatcgatAATTTACATATATACAGAAT
pFTX4K-eGFPC1-XCR1α 392-958 3'UTR	F: 5' caggatccGCATCCTTCACAAACTTTTA R: 5' gcgcatcgatAGGTACAGTATTTAATGTCT
pFTX4K-eGFPC1-XCR1α 286-637 3'UTR	F: 5' catggatccCTGCATGTTGAGATCAGGAA R: 5' gcgcatcgatGGCTAGCACTTACATTGCT
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR	F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 276-402 3'UTR	F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtAAAAGTTTGTGAAGGATGC
pFTX4K-eGFPC1-XCR1α 276-461 3'UTR	F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5'ggtactagtAAAATGTTCCTAGATACTTTG
pFTX4K-eGFPC1-XCR1α 276-579 3'UTR	F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGGGATGCAGTCACACACA

pFTX4K-eGFPC1-XCR1α 450-520 3'UTR	F: 5' ggtggatccTAGGAACATTTTACAAAT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 392-520 3'UTR	F: 5' caggatccGCATCCTTCACAAACTTTTA R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-107	First round PCR primer pairs: F: 5' GATCCCTTGCTCTTTCCACGATGTTGAGATCAGGA R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5' TCCTGATCTCAACATCGTGGAAAGAGCAAGGGATC Second round PCR primer pair: F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-499	First round PCR primer pairs: F: 5' TGAGATCAGGAAGCTTCAGTAATACTTACCCAGTGC R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5' GCACTGGGTAAGTATTACTGAAGCTTCCTGATCTCA Second round PCR primer pair: F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-gga-miR-1810	First round PCR primer pairs: F: 5'TAAATTTTATTAACTATATACACACGTTGTACATTA R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5'TAATGTACAACGTGTGTATATAGTTAATAAAATTTA Second round PCR primer pair: F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA

pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-Pub1	First round PCR primer pairs: F: 5' CTTCACAAACTTTTATATGAACTTAATCAATACATTAATTGACAAAGA R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5' TCTTTGTCAATTAATGTATTGATTAAGTTCATATAAAAGTTTGTGAAG Second round PCR primer pair: F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-XCR1α 276-520 3'UTR mut-miR-140	First round PCR primer pairs: F: 5' TACTACATTGCATGTAGTCCATCTTGTTTGTATGTT R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5' AACATACAAACAAGATGGACTACATGCAATGTAGTA Second round PCR primer pair: F: 5' catggatccCTTGCTCTTTGCTGCATGT R: 5' ggtactagtGATCAGGACAAACATACAAA
pFTX4K-eGFPC1-mut-miR-107-full length XCR1α 3'UTR	First round PCR primer pairs: F: 5' GGGTTCTTGCTCTTTCCACGATGTTGAGATCAGGAA R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: TTCCTGATCTCAACATCGTGGAAAGAGCAAGAACCC Second round PCR primer pair: F: 5' catggatccCAGAGTAACTTGAGTCTG R: 5' gcgcatcgatAGGTACAGTATTTAATGTCT

pFTX4K-eGFPC1-mut-miR-1810-full length XCR1α 3'UTR	First round PCR primer pairs: F: 5' TAAATTTTATTAACTATATACACACGTTGTACATTA R: 5' CAGGGTGGATTGAGTCCAGCC F: 5' TGGACGGCGACGTAAACGGCC R: 5'TAATGTACAACGTGTGTATATAGTTAATAAAATTTA Second round PCR primer pair: F: 5' catggatccCAGAGTAACTTGAGTCTG R: 5' gcgcatcgatAGGTACAGTATTTAATGTCT

2.1.7 Linearisation of plasmids for *in vitro* transcription

20 μg of plasmid DNA was digested with the appropriate restriction enzyme and buffer for 3 hours and then purified by phenol-chloroform extraction. 200 μl of phenol:chloroform:isoamyl alcohol (Sigma) was added to the restriction digest and the mixture was vortexed and centrifuged at 13,000 rpm for 5 minutes. The upper aqueous layer was removed and transferred to an eppendorf tube containing 20 μl of sodium acetate pH 5.2 and 600 μl of 100% ethanol. The solution was precipitated on dry ice for 10 minutes and centrifuged at 13,000 rpm for 15 minutes to obtain a DNA pellet. The pellet was washed with 80 % ethanol, dried and resuspended in 15 μl of TE. The concentration of the linearised DNA was adjusted to 1 μg/μl and linearisation was confirmed by gel electrophoresis compared to uncut plasmid DNA.

Linearisation restriction digest

20 μg plasmid DNA

20 μl 10 x BSA

20 μl NEB buffer

3 μl restriction enzyme

Up to 200 μl with autoclaved water

Incubate at 37 °C for 3 hours

3 M sodium acetate

24.6 g sodium acetate (anhydrous, Fisher)

80 ml autoclaved water

Adjust pH to 5.2 with glacial acetic acid

Up to 100 ml with autoclaved water

Filter sterilised by Stericup filtration (Millipore) and stored at room temperature.

2.1.8 Preparation of capped mRNA for *Xenopus* microinjection

40 µl of mRNA synthesis mix was added to 5 µl of linearised DNA (1 µg/µl). 5 µl of RNA polymerase (50,000 units/ml) was then added and the reaction mixture was incubated at 37 °C for 30 minutes. T7, SP6 or T3 polymerase was used as appropriate. 5 µl of 10 mM rGTP was then added and the reaction was incubated at 37 °C for a further hour. 50 µl of 50-TE, 1.25 µl of DNAse I salts, 0.5 µl RNAsin (Promega) and 2 µl of DNAse I (2.7 units/µl, Worthington) was added and incubated for 30 minutes at 37 °C to remove template DNA. The mRNA was extracted by phenol-chloroform extraction. An equal volume of phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 13,000 rpm for 5 minutes to separate the upper aqueous and lower organic phases. The aqueous phase was removed and added to an eppendorf tube containing 300 µl of 100 % ethanol and 25 µl of 4 M ammonium acetate (pH 5.6). The mRNA was precipitated on dry ice for 10 minutes and centrifuged at 13,000 rpm for 15 minutes to obtain a pellet. This was washed with 80 % ethanol, dried and resuspended in 80 µl of RNAse-free water (Ambion). The mRNA concentration was then adjusted to 1 µg/µl and stored at -80 °C.

2.1.8.1 Assessment of RNA quality by gel electrophoresis

The quality of the mRNA was assessed by confirming the integrity and size of the mRNA band by gel electrophoresis. 1 µg of RNA was added to 20 µl of formamide loading buffer and denatured by heating at 95 °C for 3 minutes. TAE gel electrophoresis

was performed as previously described, except that all gel trays and combs were washed with ethanol beforehand and 1 x TAE was RNAse-free.

mRNA synthesis mix

10 µl 10 x transcription buffer (NEB)

6.25 µl DTT (100 mM)

10 µl rATP (10 mM)

10 µl rUTP (10 mM)

10 µl rCTP (10 mM)

5 µl rGTP (1 mM)

5 µl CAP analogue (10 mM, NEB)

5 µl RNAsin (Promega)

10 µl autoclaved water

50-TE

50 mM Tris-HCl pH 7.5

1 mM EDTA

DNAse I salts

5 µl $CaCl_2$ (1 M)

10 µl $MgCl_2$ (1 M)

10 µl DTT (100 mM)

4 M ammonium acetate

30.8 g NH_4OAc (BDH laboratory supplies)

80 ml autoclaved water

Adjust pH to 5.6 with glacial acetic acid

Up to 100 ml with autoclaved water

Filter sterilised by Stericup filtration (Millipore) and stored at room temperature.

Formamide RNA loading buffer

0.25 % (w/v) bromophenol blue (Sigma)

0.25 % (w/v) xylene cyanol (Sigma)

20 mM EDTA

Made up in deionised formamide (Fisher)

2.1.9 Preparation of *in situ* hybridisation probes

Linearised DNA was used as the template for *in situ* probe transcription. RNA probes were labelled with digoxygenin-11-rUTP by including DIG-UTP in the *in vitro* transcription reaction mix. The reaction was incubated at 37 °C for 2 hours and template DNA was then removed by the addition of 48 µl of 50-TE, 2 µl of DNAse I salts, 0.5 µl of RNAsin and 1 µl of DNAse I and subsequent incubation at 37 °C for 15 minutes. RNA was then isolated by G-50 sephadex column purification (GE Healthcare) according to the manufacturer's instructions. Size exclusion on the column means that nucleic acids larger than 20 nucleotides are eluted and purified from unincorporated nucleotides. RNA concentration was estimated by Nanodrop quantification and RNA quality was confirmed by gel electrophoresis. Probes were diluted to a 10 x stock at a concentration10 ng/µl in *in situ* hybridisation buffer (section 2.3.5.1) and stored at -20 °C prior to use.

DIG-rNTP mix

5 µl rCTP (100 mM)

5 µl rGTP (100 mM)

5 µl rATP (100 mM)

3.25 µl rUTP (100 mM)

17.5 µl digoxygenenin-11-rUTP (10 mM)

164.25 µl RNAse-free water

In situ probe transcription reaction

5 µl 10 x transcription buffer (NEB)

10 µl DIG-rNTP mix

5 µl 100 mM DTT

1 µl RNAsin

2 µg linearised DNA

4 µl RNA polymerase (T7, SP6 or T3 polymerase as required, 50,000 units/ml

Up to 50 µl with RNAse-free water

2.2 *Xenopus laevis* and *Xenopus tropicalis* embryo manipulations

A list of general solutions used in *Xenopus laevis* experiments is given in Section 2.2.6.

2.2.1 *Xenopus laevis* husbandry

Laboratory-bred adult frogs were obtained from Nasco, Fort Atkinson, WI, USA and the European *Xenopus* Resource Centre, Portsmouth, UK and housed separately according to sex in a Tecniplast Xenoplus housing system. The system was set to maintain pH at 7.5, conductivity at 1100 µS and temperature at 18 °C. Frogs were fed 1

ml of amphibian diet pellets per frog biweekly. The system exchanged 20 % of its volume every day and continuously recirculated water through a pre-filter to remove solid waste and a mechanical filter to trap particulate waste. Eggs were removed by hand as necessary. Water quality was monitored weekly by measuring levels of nitrates, nitrites, and ammonia using the Palintest water testing kit.

2.2.2 Obtaining *Xenopus laevis* oocytes

Female frogs were injected in the dorsal lymph sac with 50 units of PMSG (Intervet) 3-5 days before induction of ovulation and subsequently injected with 500 units of HCG (Sigma) the afternoon prior to egg collection. Stimulated frogs were then massaged to release eggs from the cloaca into a petri dish. Excess water was removed from the oocytes before storage at 16 °C prior to *in vitro* fertilisation.

2.2.3 *In vitro* fertilisation of *Xenopus laevis* oocytes

An adult *X. laevis* male was killed according to Schedule I protocol by immersion in 0.2 % (w/v) MS-222 for 30 minutes. Death was confirmed by severing the spinal chord. Testes were isolated by dissection and placed in Leibovitz's L-15 medium (Gibco). Testes were crushed in 500 μl of high salt 1 x MBS using a micropestle and the solution was mixed by repeated pipetting and made up to 1.5 ml with high salt 1 x MBS. This solution was distributed evenly over the oocytes in the petri dish and allowed to incubate for 5 minutes. The petri dish was then flooded with 1/10 x NAM and left at room temperature for 30 minutes. During this time, the fertilised eggs undergo cortical rotation within the vitelline membrane so that the animal hemisphere faces upwards.

2.2.4 Embryo manipulations

2.2.4.1 De-jellying and sorting

Embryos were prepared for subsequent manipulation by removal of the jelly coat. Fertilised eggs were transferred to a beaker containing 2 % (w/v) cysteine (pH 7.8) and gently agitated until the fertilised eggs began to pack together, which indicates de-jellying has occurred. The fertilised eggs were washed four times in 1/10 x NAM to remove the cysteine solution and transferred to a clean petri dish. Eggs were sorted for fertilisation by selecting those resistant to deformation with a hair loop tool.

2.2.4.2 Microinjection

Fertilised embryos were injected using a Narishige NM-300 microinjector in 3 % (w/v) Ficoll solution at various developmental stages depending upon the experiment in question. Morpholinos and mRNA for overexpression were injected at the 1-cell stage or the 4-cell stage for GFP-3'UTR reporter mRNA experiments where the spread of the injected mRNA was critical (see Figure 3.12). Embryos were injected in the most vegetal four blastomeres at the 32-cell stage for targeting of the prospective endoderm (Figure 3.3). An injection volume of 8 nl was used in all cases. Morpholino doses varied between 20 ng and 40 ng and individual mRNA doses between 200 pg and 700 pg per embryo. After microinjection, embryos were transferred to 1/10 x NAM and sorted for survival. A list of antisense oligos used for loss of function microinjection experiments is given in Table 2.1 and Table 2.2.

2.2.4.3 Microdissection

Embryos were dissected into animal and vegetal hemispheres at stage 9, or into animal caps, marginal zone and vegetal poles at stage 10 for subsequent Western blot analysis or total RNA extraction. Dissections were carried out in petri dishes coated with 1 % agarose and containing 1 x NAM supplemented with 0.1 % (w/v) BSA to avoid sticking of the embryo to the forceps during dissection.

2.2.5 Fixation

Embryos of the desired stage were fixed for downstream analysis by confocal microscopy or *in situ* hybridisation by removing the embryos from 1/10 x NAM and adding them to a Wheaton vial (Sigma) containing 4 % paraformaldehyde in RNAse-free PBS and incubating for either 2 hours at room temperature or overnight at 4 °C. For *in situ* hybridisation, embryos were transferred from fixing solution to methanol and stored at -20 °C until they were processed.

2.2.6 General *Xenopus* solutions

All solutions were prepared using MilliQ-filtered pure, deionised water.

10 x Normal Amphibian Media (NAM)

32.5 g NaCl

0.75 g KCl

1.2 g $CaNO_3$

1.25 g $MgSO_4$

1 ml 0.5 M EDTA

Up to 500 ml with water

10 x high salt Modified Barth's Solution (MBS)

51.3 g NaCl

0.75 g KCl

2 g $MgSO_4$

2 g $NaHCO_3$

Up to 1 L with water and adjust pH to 7.6

1 x NAM

100 ml 10 x NAM salts

2 ml 1M HEPES-NaOH (pH 7.6)

10 ml 0.1 M $NaHCO_3$

Up to 1 L with water

1 x high salt MBS

50 ml 10 x high salt MBS

2 ml 5 M NaCl

350 µl 1 M $CaCl_2$

Up to 500 ml with water

1/10 x NAM

20 ml 10 x NAM salts

4 ml 1 M HEPES-NaOH (pH 7.6)

Up to 2 L with water

2 % Cysteine

6 g Cysteine (Sigma)

300 ml 1/10 x NAM

Adjust pH to 7.8 with NaOH

0.2% MS-222

4 g of Ethyl 3-aminobenzoate methanesulfonate (MS-222)

2 L of water

PMSG

20 ml of the accompanying solvent was added to a vial of PMSG containing 5000 units of hormone to make a stock of 0.25 units/ml. This was aliquoted at stored at -20 °C. 0.2 ml of this solution was injected (50 units).

HCG

6 ml of autoclaved water was added to a vial of HCG containing 10,000 units of hormone to make a stock of 1667 units/ml. 0.3 ml of this solution was injected (500 units).

2.2.7 *Xenopus tropicalis in vitro* fertilisation

Wild type *Xenopus tropicalis* embryos were obtained in the Zimmerman laboratory at the National Institute for Medical Research. All animal handling was carried out by Zimmerman lab members and I performed the *in vitro* fertilisations, sorting, staging and microdissection of embryos. A list of *Xenopus tropicalis* solutions used is provided in Section 2.2.9. Petri dishes were coated with 1/20 x MMR with 0.1% BSA added before use to prevent the embryos sticking to the plastic. Female frogs were boosted the night before a frog experiment with 10 units of HCG and injected with 200 units of HCG four hours prior to squeezing. Testes were obtained from 2-3 male frogs as described in section 2.2.3 and were stored in 10 % calf serum (v/v) in Leibovitz's L-15 medium at 4 °C prior to use. Eggs were obtained from female frogs by holding the frog above a petri dish. A mixture of eggs from three females was collected. Testes were crushed 10 % calf serum/L-15 medium and spread evenly over the eggs. A few minutes were allowed for sperm attachment, then the petri dish was flooded with 1/20 x MMR and left at room temperature for 30 minutes. Unlike *X. laevis*, fertilised *X. tropicalis* embryos do not undergo cortical rotation.

2.2.8 *Xenopus tropicalis* embryo manipulations

2.2.8.1 De-jellying and sorting

Embryos were de-jellied by swirling fertilised embryos in 2.2 % cysteine solution (pH 7.8) for 7 minutes and washing four times in 1/20 x MMR with 0.1 % BSA. Sorting for fertilisation was done at the 2-cell stage, as unfertilised embryos do not cleave. Uninjected embryos were incubated at 25 °C in 1/20 x MMR plus 0.1 % BSA for staging, or transferred to injection dishes containing 3% Ficoll for microinjection.

2.2.8.2 Microinjection

Microinjection of fertilised embryos was performed at the 2-cell stage, with half the total morpholino dose injected into each blastomere. An injection volume of 2 nl and a maximum total dose of 15 ng of morpholino was used. Morpholinos were coinjected with fluorescein dextran as a tracer. All *X. tropicalis* morpholino injections were performed by Holly Ironfield in the Zimmerman laboratory. After injection, embryos were allowed to recover in Ficoll for 2 hours at 25 °C, then transferred to 1/20 x MMR

plus 0.1 % BSA for staging. Development of *X. tropicalis* embryos can be accelerated by incubating at 25-28 °C or slowed by incubating at 22 °C.

2.2.8.3 Microdissection

Microdissection of *X. tropicalis* embryos into animal and vegetal halves was performed at stage 10 exactly as described in section 2.2.4.3.

2.2.9 *Xenopus tropicalis* solutions

All solutions were made with MilliQ-filtered pure deionised water.

1 x Marc's Modified Ringers (MMR)

0.1 M NaCl

2 mM KCl

1 mM $MgSO_4$

2 mM $CaC1_2$

5 mM HEPES pH 7.8

0.1 mM EDTA

1/20 x MMR + gentamycin

50 ml 1 x MMR

1 ml gentamycin (100 mg/ml)

Up to 1L with water

1/20 x NAM + 0.1% (w/v) BSA

1 L 1/20 x NAM

1 g BSA (Sigma)

2.2 % Cysteine

2.2 g Cysteine.HCl (Sigma)

100 ml 1/20 x MMR

pH to 7.8 with NaOH

3% Ficoll

3 g Ficoll (Sigma)

100 ml 1/20 x MMR

10 x Fluorescein dextran

250 mg fluorescein dextran (Sigma)

50 ml autoclaved water

Morpholino	Sequence	Description
MoC	5'CCTCTTACCTCAGTTACAATTTATA	Control morpholino against human β-globin
Mo-xla-miR-427	5'ACGCCCAAAACAGAAAGCACTTTCTCATCAT	Multi-blocker microRNA morpholino designed to inhibit processing of mature xla-miR-427 from its hairpin precursor by blocking Dicer site (Rosa et al., 2009).
MoDicer *(X. laevis)*	5'GAGTCATGAGCTGAAGGCCTGCCAT	Translation blocking *X. laevis* Dicer morpholino over the ATG (Gessert et al., 2010).

TP-Pumilio	5'TGTACATATATACA GAATGCACTGG	Antisense morpholino used as a target protector to block the Pumilio sites (PBE1-2) in the *XCR1α* 3'UTR.
Mo-miR-F *(X. laevis and X. tropicalis)*	5'AGCCGAACGAGAA GCACTTTCTCAT	MicroRNA morpholino against novel *Xenopus* microRNA, miR-F. Morpholino is targeted to the guide strand and overlaps the Drosha site by 2 nt, which should prevent some processing of the primary miRNA transcript to a hairpin precursor in the nucleus.
MoDicer *(X. tropicalis)*	5'GATGTCCAGTAAT CATACCTGATAT	Splice-blocking *X. tropicalis* morpholino over the exon1-intron1 boundary of Dicer prevents splicing out of intron1 that introduces a stop codon to produce a truncated protein.

Table 2.1 Morpholinos used in *Xenopus* microinjection

Antagomir	Sequence	Description
Control antagomir	Undisclosed	Control RNA oligo (Ambion, catalogue number 17010)
α-miR-1810 antagomir	5'AGCUCACGUUCCCUAUUAGU	Antisense RNA oligo to the mature gga-miR-1810 miRNA sequence (Ambion, catalogue number 17010)

Table 2.2 Antagomirs used in *X. laevis* microinjection

2.3 RNA Analysis

All RNA procedures were carried out taking precautions to avoid RNAse contamination and degradation of the RNA sample, including using filter tips, RNAse-free water, autoclaved solutions and regular changing of gloves. All RNA samples were centrifuged at 4 °C and kept on ice during the experiment.

2.3.1 Total RNA extraction from *Xenopus* embryos

Total RNA was extracted from embryos snap-frozen on dry ice and stored at -80 °C by the addition of 1ml of Trizol reagent (Invitrogen). Embryos were homogenised by repeated syringing through a 21G needle and light vortexing. The homogenate was then centrifuged at 13,000 rpm for 10 minutes at 4 °C and the resulting supernatant was transferred to a new eppendorf tube. 200 μl of chloroform was added and the mixture was vortexed and spun at 13,000 rpm for 15 minutes at 4 °C to separate the aqueous phase containing the RNA, and the organic phase. The aqueous layer was removed and transferred to an eppendorf tube containing 500 μl of isopropanol and vortexed. To precipitate nucleic acid, the samples were incubated on dry ice for 10 minutes and centrifuged at 13,000 rpm to obtain a pellet. Following removal of the supernatant, the pellet was washed in 80% ethanol and resuspended in 50 μl of RNAse-free water.

DNAse I digestion of any contaminating genomic DNA was then performed by addition of 95 μl of 50-TE, 0.5 μl of RNAsin (Promega) and 1 μl (2.7 units) of DNAse I (Worthington) and incubation at 37°C for 30 minutes. The RNA was extracted by addition of 150 μl of phenol:chloroform:isoamyl alcohol (Sigma), vortexing and centrifugation at 13,000 rpm for 5 minutes. The aqueous layer containing the RNA was added to an eppendorf tube containing 300 μl 100% ethanol and 25 μl of 4 M ammonium acetate (pH 5.6) and incubated on dry ice for 10 minutes. An RNA pellet was obtained by centrifugation at 13,000 rpm for 15 minutes. After washing with 80 % ethanol, the RNA pellet was resuspended in 240 μl of RNAse-free water and quantified using the Nanodrop. The RNA quality was assessed by gel electrophoresis as described in Section 2.1.8.1. 250 μl of isopropanol was added to the RNA for storage as a slurry at

-80°C. Total RNA samples from mammalian cells were obtained from members of the Developmental Signalling Laboratory.

2.3.2 Quantitative PCR

0.5 μg of total RNA was used as the input in the reverse transcription cDNA synthesis reaction using the AffinityScript QPCR cDNA synthesis kit (Stratagene) according to the manufacturer's instructions. The reverse transcription reaction was primed with random hexamers and a negative control cDNA synthesis reaction without reverse transcriptase enzyme was included (no RT control) to assess the level of genomic DNA contamination in the qPCR reaction. The 20 μl cDNA reaction was diluted by adding 80 μl of RNAse-free water and stored on ice prior to use.

cDNA synthesis reaction

10.0 μl of first strand master mix (2 x)

3.0 μl of random primers (0.1 μg/μl)

1.0 μl of AffinityScript reverse transcriptase

0.5 μg of RNA

Up to 20 μl with RNAse-free water

25 °C 5 minutes

42 °C 15 minutes

55 °C 15 minutes

95 °C 5 minutes

4 °C ∞

A 15 μl mastermix containing EXPRESS SYBR GreenER qPCR Supermix (Invitrogen) and qPCR primers was prepared. This was pipetted into a 96 well optical

plate (Applied Biosystems) using an electronic pipette and 5 μl of cDNA was added. Triplicate repeats of samples were pipetted. The plate was sealed and centrifuged for 2 minutes at 2000 rpm. The qPCR reaction was performed using an ABI 7500 Fast Real-Time PCR machine. EXPRESS SYBR GreenER qPCR Supermix contains Taq polymerase, dNTPs and SYBR GreenER dye. SYBR GreenER binds double-stranded DNA and emits green light in proportion to the amount of dsDNA bound when excited by blue light. 40 cycles of PCR amplification were performed in the ABI 7500 Fast machine and fluorescence was measured at each cycle. A Ct value was recorded for each reaction. This is the cycle number required for the fluorescence to reach a threshold level, and is inversely proportional to the abundance of the PCR product. Ct values for each gene of interest were normalised to the Ct value of ODC, which has a constant mRNA level throughout *Xenopus* development (Paris et al., 1988). mRNA levels for the gene of interest were calculated relative to ODC and standard deviations of the triplicate repeats were plotted as error bars.

$$\text{mRNA level relative to ODC} = 2^{\,Ct(ODC)-\,Ct(\text{Gene of interest})}$$

A melt curve was also run to check the specificity of the PCR amplification. In a melt curve, fluorescence is measured against temperature. Heating DNA causes dissociation of the two strands, and a reduction in fluorescence as SYBR GreenER binding decreases. The temperature required to dissociate or melt the DNA varies with its length and so multiple peaks in melt curves can indicate the presence of multiple PCR products.

qPCR reaction mix
10.0 μl EXPRESS SYBR GreenER qPCR Supermix (Invitrogen)
0.4 μl Forward primer (10 μM)
0.4 μl Reverse primer (10 μM)
4.2 μl water
5.0 μl diluted cDNA

qPCR primer pairs	
X. tropicalis Dicer Exon1-Exon2	F: 5' CGGAAATATCAGGTGGAACTAC R: 5'AGTCCCCACGGATCTGGTAT
X. tropicalis Dicer Exon1-Intron1	F: 5' CTGGCAACAGGAGGCTATTC R: 5' CCACTGCAATGATTCTGTTCTT
X. tropicalis ODC	F: 5' GGCCACACTGGCAACTCATGC R: 5' CGGTGTGCGCTCAGTTCTGGT

2.3.3 RNAse protection assay

2.3.3.1 RNA probe synthesis and purification

A radiolabelled antisense RNA probe against the mRNA target of interest was synthesised by *in vitro* transcription and gel purified. Firstly, 0.5 μg of linearised template DNA was incubated with probe synthesis mix, radiolabelled rUTP and either T7 or SP6 polymerase at 37 °C for 1 hour. Template DNA was then removed by DNAse I digestion. 1 μl of DNAse I (2.7 units) was added to the probe synthesis reaction and incubated at 37 °C for 30 minutes. 10 μl of formamide RNA loading buffer was added, and the sample was denatured by heating at 95 °C for 3 minutes and loaded onto a 6 % denaturing acrylamide gel. The gel was run with 0.5 x TBE running buffer at 13 W until the bromophenol blue had run off. The gel was covered in saran wrap and the position of the probe band on the gel was determined by a 1-minute exposure of the gel on x-ray film. The probe band was excised and the gel slice was incubated in 350 μl of probe elution buffer at 50 °C for 2 hours. The probe elution solution was then

transferred to a new eppendorf tube and 350 µl of phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 13,000 rpm for 5 minutes. The upper, aqueous layer was transferred to an eppendorf tube containing 3 µl of 1 M $MgCl_2$, 5 µl of yeast tRNA (10 µg/µl) and 900 µl of 100 % ethanol and the RNA was precipitated on dry ice for 10 minutes. This was followed by centrifugation at 13,000 rpm for 20 minutes to obtain an RNA pellet which was washed in 80 % ethanol and dissolved in 20 µl of RNAse-free water. 1 µl of the RNA probe was counted in a scintillation counter and the probe concentration was adjusted to 2 x 10^5 cpm/µl.

Probe synthesis reaction

1.75 µl probe synthesis mix (see below)

0.5 µl linearised DNA (1µg/µl)

0.5 µl RNA polymerase (25 units)

2.5 µ α-^{32}P- rUTP (800 Ci/mmol, Perkin Elmer)

Probe synthesis mix

4 µl 10 x transcription buffer (NEB)

2 µl rATP (10 mM)

2 µl rGTP (10 mM)

2 µl rCTP (10 mM)

2 µl RNAsin (Promega)

2 µl DTT (100 mM)

6 % denaturing gel mix (30 ml)

1.5 ml 10 x TBE

1.5 ml water

19.8 ml Sequagel diluent (National diagnostics)

7.2 ml Sequagel concentrate (National diagnostics)

60 µl TEMED

75 µl 20 % (w/v) APS

2.3.3.2 Probe hybridisation to RNA

Between 10 and 20 µg of *Xenopus* total RNA was spun out from a slurry (see section 2.3.1). The pellet was washed in 80 % ethanol and resuspended in 30 µl of hybridisation buffer. 1 µl of each antisense radiolabelled probe was added and the sample was incubated at 85 °C for 5 minutes to allow denaturation of the RNA. Hybridisation was carried out overnight at 45 °C.

2.3.3.3 RNAse digestion

The hybridised RNA was added to 350 µl of RNAsing solution and 0.5 µl RNAse T1 (20 units/µl, Calbiochem). The RNAse digestion reaction was incubated at 37 °C for 30 minutes. Next, 10 µl of 20 % (w/v) SDS and 5 µl of Proteinase K (10 µg/µl) was added to and the samples were incubated at 37 °C for 30 minutes to remove the RNAse T1 enzyme. The samples were then extracted by addition of 350 µl of phenol:chloroform:isoamyl alcohol (Sigma), vortexed and centrifuged at 13,000 rpm for 5 minutes. The aqueous layer was then added to 500 µl of isopropanol with 15 µg of yeast tRNA added. The RNA was precipitated on dry ice for 10 minutes and centrifuged at 13,000 rpm at for 20 minutes to obtain an RNA pellet. The pellets were washed with 80 % ethanol, dried and resuspended by adding 1 µl of RNAse-free water and 5 µl of formamide RNA loading buffer. Samples were denatured by heating at 95 °C for 3 minutes.

2.3.3.4 Detection of protected mRNA fragments

Denatured RNA samples and a radiolabelled low molecular weight ladder were loaded onto a 6 % denaturing gel and run for 1 hour and 15 minutes at 13 W. The gel was then fixed by immersion in 10 % methanol/10 % acetic acid in water for 15 minutes and dried on a gel dryer. The dried gel was then exposed overnight at -80 °C to obtain the signal for the protected mRNA fragment.

2.3.3.5 RNAse protection assay solutions

Probe elution buffer

0.5 M ammonium acetate

0.1 % (w/v) SDS

1 mM EDTA

Made up in RNAse-free water

1 x Hybridisation buffer

400 mM NaCl

40 mM PIPES pH 6.4

1 mM EDTA

80 % (v/v) formamide

RNAsing solution

10 mM Tris pH 7.5

5 mM EDTA

300 mM NaCl

Made up in RNAse-free water

Radioactive marker preparation

1 µl (0.5 µg) low molecular weight marker (NEB)

1 µl NEB buffer 2

1 µl nucleotide mix (330 µM dATP, dTTP, dGTP)

1 µl α-^{32}P- dCTP (3000 Ci/mmol)

6 µl water

1 µl Klenow (50 units/µl NEB)

The reaction was incubated at room temperature for 10 minutes and the radiolabelled marker was diluted by addition of 40 µl of TE and 50 µl of formamide dyes. 1 µl of a 1/10 dilution of the marker was used.

2.3.4 Northern blot analysis of transcripts

20 µg to 30 µg of *Xenopus* embryo total RNA was analysed for transcript levels by Northern blot. RNA was incubated with an equal volume of denaturing solution at 50 °C for 1 hour. 1/5 volumes of 6 x gel loading buffer was added to the denatured RNA and the samples were loaded onto a 1.5 % agarose gel made with sodium phosphate running buffer (pH 6.8). The gel was run at 4 V/cm according to the length of the gel until the bromophenol blue had run 10 cm. After electrophoresis, unused areas of the RNA gel were trimmed away and the bottom left corner was cut to orient the gel. The transfer apparatus was then assembled. The nylon transfer membrane (Hybond N+, Amersham) was prepared by cutting to the same size as the gel, pre-wetting in water and soaking in 20 x SSC (Gibco) for 5 minutes. In an autoclaved glass baking dish, a sterile tissue culture flask was used as a support and covered with a piece of Whatman paper. The dish was then filled with 20 x SSC and the wet Whatman paper was rolled to remove any air bubbles. The RNA gel was then placed face down on the wet Whatman paper. The gel was then surrounded with saran wrap to prevent short-circuiting of the transfer. The nylon membrane was placed on top of the gel and a stack of paper towels

were placed on top of the membrane and held in position by a glass plate and a 500 g weight. Transfer of the RNA onto the nylon membrane took place overnight.

The transfer apparatus was disassembled and the nylon membrane was washed in 6 x SSC for 5 minutes to remove any agarose stuck to it. The membrane was dried between two pieces of Whatman paper and the RNA was crosslinked to the membrane using a UV-crosslinker (Stratagene) set to 1 minute at 120,000 $\mu J/cm^2$ and repeated on both sides of the membrane. Successful transfer of RNA was confirmed by staining with methylene blue and destaining with water to reveal ribosomal RNA bands. The membrane was scanned at this point to obtain a loading control. The membrane can be stored dry at 4 °C until use or incubated with 50 ml of hybridisation buffer at 65 °C in a hybridisation oven for 1-4 hours. In the meantime, a radiolabelled antisense RNA probe was synthesised and purified as described in Section 2.3.3.1. The RNA probe was added to 50 ml of fresh hybridisation buffer and hybridised to the membrane overnight by rotating incubation in the hybridisation oven at 65 °C.

The membrane was then washed by exchanging the hybridisation buffer for washing solution. Four 5-minute room temperature washes in washing solution 1 followed by two washes for 10 minutes at 60 °C in washing solution 2 were carried out. The damp membrane was covered in saran wrap and placed in a cassette with an intensifying screen and exposed for 6 hours at -80 °C, then longer as necessary to develop the signal.

2.3.4.1 Northern blot solutions

1 M Sodium phosphate buffer pH 6.8

46.3 ml 1 M Na_2HPO4

53.7 ml 1 M NaH_2PO4

6 M Glyoxal (deionised)

20 ml of glyoxal was deionised by stirring with 20 g of AG501-X8 ion-exchange resin (Biorad) until the resin was colourless. A further 4 g of resin was added and stirred for 30 minutes. This was repeated as needed until the pH reached 5.5. Deionised glyoxal was aliquoted and stored at -20 °C.

Denaturing solution

66 µl of 6 M deionised glyoxal

1 µl 1 M sodium phosphate buffer pH 6.8

1 µl 0.5 M EDTA

133 µl DMSO

6 x Loading buffer

0.25 % (w/v) bromophenol blue

0.25 % (w/v) xylene cyanol

30 % (w/v) glycerol in water

1 x Running buffer

10 mM sodium phosphate buffer pH 6.8

Methylene blue

0.5 M sodium acetate pH 4.5

0.04 % methylene blue

Hybridisation buffer

50 % (v/v) deionised formamide (Fisher)

5 x SSPE (Sigma)

5 x Denhardt's solution

0.5 % (w/v) sodium pyrophosphate

0.1 μg/ml denatured, sheared herring sperm DNA (Sigma)

10 % (w/v) dextran sulfate

2 % (w/v) SDS

Made up in autoclaved water

100 x Denhardt's solution

2 % (w/v) BSA

2 % (w/v) polyvinylpyrrolidone (PVP-40, Sigma)

2 % (w/v) Ficoll (Sigma)

Made up in autoclaved water, aliqoted and stored at -20 °C.

Washing solution 1

2 x SSPE (Sigma)

0.1 % (w/v) SDS

Made up in autoclaved water

Washing solution 2

0.1 x SSPE (Sigma)

0.1 % (w/v) SDS

Made up in autoclaved water

2.3.5 *Xenopus in situ* hybridisation

In situ hybridisation for mRNA of developmental markers of interest was performed on embryos fixed at the appropriate stage and stored in methanol. *In situ* hybridisations were carried out in Wheaton vials or homemade baskets. Washes were at room temperature unless otherwise stated. Embryos were rehydrated by 5 minute washes in 75 % and 50 % ethanol in water, followed by 25 % ethanol in 75 % PBS-Tween. After two 5-minute washes in PBS-Tween, embryos were bleached by incubation in bleaching solution under bright light for 10 minutes. Bleaching solution was removed by two 5-minute washes in PBS-Tween. To ensure embryo integrity after bleaching, embryos were re-fixed for 20 minutes in 4 % paraformaldehyde in PBS-Tween. Embryos were then equilibrated in hybridisation buffer at 62 °C for between 2 and 3 hours. This solution was then replaced by a 1x dilution of antisense probe in hybridisation buffer and incubated overnight 62 °C to allow hybridisation of the probe to the target mRNA.

After recovery of the probe for later reuse, several 10-minute washing steps were performed at 62 °C to reduce non-specific probe binding. Embryos were washed in hybridisation buffer, followed by 25 % formamide, 2 x SSC and 0.1 % Tween, then 12.5 % formamide, 2 x SSC and 0.1 % Tween, then 2 x SSC and 0.1 % Tween. A final 30-minute wash at 62 °C in 0.2 x SSC and 0.1% Tween was performed. Embryos were then washed three times at room temperature in PBS-Tween for 5 minutes. Following equilibration in maleic acid buffer by washing twice for 5 minutes, embryos were incubated in blocking solution for 2 hours. Blocking solution was then replaced by anti-DIG antibody conjugated to alkaline phosphatase (Roche) at a dilution of 1 in 2000 in blocking solution and incubated overnight at 4 °C.

Post-antibody washes were performed in maleic acid buffer (MAB) with 0.1 % Tween for 6 hours at room temperature and subsequently overnight at 4 °C. Embryos were then washed twice for 10 minutes in alkaline phosphatase buffer to adjust the pH to 9.5 for the alkaline phosphatase reaction. The embryos were transferred to 24-well plates and staining was developed by addition of 1 ml of BM Purple reagent (Roche). Development time was varied as required at room temperature for several hours, or

overnight at 4 °C. Once a satisfactory staining was achieved, embryos were fixed in 4 % paraformaldehyde in PBS and images were acquired on a Leica dissecting microscope.

2.3.5.1 In situ solutions

PBS-Tween

1 x PBS

0.1 % (v/v) Tween-20 (Sigma)

Bleaching solution (1ml)

600 µl RNAse-free water

325 µl 30 % (v/v) hydrogen peroxide (Sigma)

50 µl deionised formamide (Fisher)

25 µl 20 x SSC (Gibco)

Up to 1 ml with RNAse-free water

Hybridisation buffer

50 % (v/v) deionised formamide

5 x SSC

1 mg/ml yeast RNA (torula, Sigma)

0.1mg/ml heparin (Sigma)

1 x Denhardt's solution

0.1 % Tween-20

10 mM EDTA

Made up in RNAse-free water

5 x Maleic Acid Buffer

43.83 g NaCl

58.03 g Maleic acid

Adjust pH to 7.5 with sodium hydroxide

Up to 1L with RNAse-free water

Blocking solution

2 g blocking reagent (Roche)

20 ml 5 x MAB

80 ml RNAse-free water

Alkaline Phosphatase Buffer

100 mM Tris pH 9.5

50 mM $MgCl_2$

100 mM NaCl

0.1 % Tween

2.4 MicroRNA Analysis

2.4.1 MicroRNA Northern blot analysis

2.4.1.1 RNA separation and transfer

30 μg of *Xenopus* embryo total RNA was spun out from an isopropanol slurry and resuspended in 30 μl of formamide RNA loading buffer and denatured by heating at 95 °C for 3 minutes. The samples and a radiolabelled small RNA ladder (Decade markers, Ambion, prepared according to the manufacturer's instructions) were loaded onto a 15% denaturing gel. The gel was run using 1 x TBE running buffer at 10 W for 10 minutes, and then 30 W for 35 minutes. To visualise the separation of the RNA and obtain a loading control, the gel was stained in 0.5 x TBE with 10 μl of 1 % ethidium

bromide solution added. The gel was then wrapped in saran and the RNA was transferred to a neutral (Hybond N, Amersham) or positively charged (Hybond N+, Amersham) nylon membrane by semi-dry blotting. The transfer was assembled by wetting 3 pieces of Whatman paper cut to the same size as the gel in 0.5 x TBE and placing them on the semi-dry transfer apparatus. Next, the nylon membrane was placed on top of the Whatman paper. The membrane was pre-wetted in water and equilibrated in 0.5 x TBE. Next the gel was placed on top of the membrane and 3 more pieces of wet Whatman paper were placed on top of the gel. The transfer sandwich was rolled to remove air bubbles and transfer was performed with a current of 2 mA/cm^2 for 1.5 hours with a maximum voltage of 25 V.

After transfer of RNA, the membrane was dried and cross-linked either by UV cross-linking or chemical cross-linking. For UV cross-linking, both sides of the membrane were cross-linked for 1 minute at 120,000 μJ/cm^2 in a UV crosslinker (Stratagene). Chemical cross-linking improves small RNA detection compared to UV cross-linking by approximately 5-fold (Pall and Hamilton, 2008) and so much less RNA can be used (5-10 μg is sufficient for signal detection, see Figure 4.21). Chemical cross-linking requires transfer onto a neutral nylon membrane. For chemical cross-linking, the damp membrane is placed face up onto 2 Whatman papers saturated with freshly prepared EDC cross-linking solution. The membrane and saturated papers are wrapped in saran and incubated for 1 hour at 60 °C. After cross-linking, the membrane is washed in RNAse-free water to remove residual cross-linking solution. After cross-linking by UV or chemical methods, the membrane can be stored dry at 4 °C or placed directly into a 50 ml falcon tube containing 15 ml of hybridisation buffer and equilibrated at 39 °C in a hybridisation oven for 1 hour.

2.4.1.2 microRNA probe labelling

Antisense DNA oligos for each microRNA of interest were obtained from Sigma. Oligos were end-labelled by the addition of a 5' radiolabelled phosphate in an *in vitro* kinase reaction using γ-^{32}P-ATP. The reaction was incubated at 37 °C for 1 hour. Labelled probe was purified from unincorporated nucleotides by applying the reaction to a G-25 sephadex column (GE healthcare).

Probe labelling by in vitro kinase reaction

1 μl antisense DNA oligo (20 μM)

5 μl γ- ^{32}P-ATP (6000 Ci/mmol)

2 μl 10 x T4 polynucleotide kinase buffer

1 μl T4 polynucleotide kinase (20,000 units/ml, NEB)

11 μl water

2.4.1.3 Probe hybridisation, washing and signal detection

The pre-hybridisation solution was poured off the membrane and replaced with 15 ml of hybridisation solution containing 20 μl of purified probe. The membrane was incubated overnight under rotation in a hybridisation oven at 39 °C. The membrane was then washed twice for 10 minutes with wash buffer at 39 °C. The membrane was covered in saran wrap and exposed overnight at -80 °C or longer as needed to obtain the microRNA signal.

2.4.1.4 MicroRNA Northern blot solutions

15 % denaturing gel mix (200 ml)

71.2 ml 40 % acrylamide (Fisher)

3.76 ml 2 % bis (Fisher)

84 g urea

50 ml 10 x TBE

75 ml RNAse-free water

50 μl of TEMED and 150 μl of 20 % APS (w/v) was added to 50 ml of denaturing gel mix and 30 minutes was allowed for the gel to polymerise.

Hybridisation buffer (30 ml)

9 ml 20 x SSC

0.6 ml 1 M Na_2HPO_4 pH 7

0.75 ml 20 % (w/v) SDS

0.3 ml 100 x Denhardt's

0.3 ml denatured, sheared herring sperm DNA (1 mg/ml Sigma)

19.05 ml RNAse-free water

Wash buffer

60 ml 20 x SSC

10 ml 20 % (w/v) SDS

130 ml RNAse-free water

EDC chemical cross-linking solution

0.16 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (Sigma)

0.13 M 1-methylimidazole pH 8.0 (Sigma)

Made up in RNAse-free water

24 ml of EDC cross-linking solution is sufficient to saturate 320 cm^2 of Whatman paper

2.4.2 MicroRNA qPCR

The microRNA qPCR procedure is described in Figure 4.19. 5 μg of *Xenopus* embryo or mammalian cell line RNA was polyadenylated by incubation with *E. coli* poly(A) polymerase (NEB) for 1 hour at 37 °C. The polyadenylated total RNA was extracted using phenol-chloroform organic extraction. 80 μl of RNAse-free water and 100 μl of phenol:chloroform:isoamyl alcohol (Sigma) was added to the polyadenylation reaction, vortexed and centrifuged at 13,000 rpm for 5 minutes. The upper aqueous layer was

added to an eppendorf tube containing 300 μl of 100 % ethanol and 25 μl of 4 M ammonium acetate (pH 5.6) and precipitated on dry ice for 10 minutes. The sample was then centrifuged at 13,000 rpm at 4 °C for 15 minutes to obtain an RNA pellet. The pellet was washed with 80 % ethanol, dried, and resuspended in 4 μl of RNAse-free water and the concentration of the polyadenylated RNA was adjusted to 1 μg/μl. 3 μg of polyadenylated RNA was used as the template for cDNA synthesis primed with a poly(T) adapter to reverse transcribe all polyadenylated RNA. The 20 μl cDNA reaction was diluted by addition of 80 μl of water. 5 μl of diluted cDNA was used in the qPCR reaction. The microRNA qPCR reaction was performed as described in Section 2.3.2, except that a universal reverse primer complementary to the poly(T) adapter sequence is used in all microRNA qPCR reactions. The forward primer is identical to the mature microRNA sequence, except that U is replaced by T. MicroRNA levels relative to ODC or U6 were calculated and error bars showing standard deviations of the triplicate repeats were plotted.

Polyadenylation reaction

2 μl 10 x poly(A) polymerase reaction buffer (NEB)

2 μl dATP (10 mM)

5 μg total RNA

1 μl poly(A) polymerase (5 units, NEB)

Up to 20 μl with RNAse-free water

cDNA synthesis reaction

10.0 µl of first strand master mix (2 x)

3.0 µl of poly(T) adapter (0.1 µg/µl)

1.0 µl of AffinityScript reverse transcriptase

3.0 µl of polyadenylated RNA (1 µg/µl)

3 µl RNAse-free water

25 °C 5 minutes

42 °C 15 minutes

55 °C 15 minutes

95 °C 5 minutes

4 °C ∞

miRNA-qPCR primers	
poly(T) adapter for cDNA synthesis	5’ GCGAGCACAGAATTAATACGACTCACT ATAGGTTTTTTTTTTTVN
MicroRNA qPCR reverse primer	5’ GCGAGCACAGAATTAATACGACTCAC
MicroRNA qPCR forward primers	
miR-F	5’ GAGAAAGTGCTTCTCGTTCGGCTGA
gga-miR-1810	5’ ACTAATAGGGAACGTGAGCT
xtr-miR-427	5’ GAAAGTGCTTTCTGTTTTGGGCG
PIR 32037	5’GCAGGACGGTGGCCATGGAAGTCGGAA TC
xtr-miR-A	5’ AGCAAATCTGTTGGTTTGTACAAAC
xtr-miR-C	5’ TGCGCTCGGTGGATAGAAGACGTGA
xla-miR-427	5’ AAAGTGCTTTCTGTTTTGGGCG
xla-miR-16c	5’ TAGCAGCACGTAAATACTGGAG
cfa-miR-30e	5’ CTTTCAGTCGGATGTTTACAGC

xtr-let-7c	5’ TGAGGTAGTAGGTTGTATGGTT
xtr-let-7f	5’ TGAGGTAGTAGATTGTATAGTT
xtr-miR-10a	5’ TACCCTGTAGATCCGAATTTGTG
xtr-miR-16c	5’ TAGCAGCACGTAAATACTGGAG
xtr-miR-101a	5’ TACAGTACTGTGATAACTGAAG
xtr-miR-148a	5’ TCAGTGCACTACAGAACTTTGT
xtr-miR-182	5’ TTTGGCAATGGTAGAACTCACA
xtr-miR-184	5’ TGGACGGAGAACTGATAAGGCT
xtr-miR-206	5’ TGGAATGTAAGGAAGTGTGTGG
bta-miR-140	5’ TACCACAGGGTAGAACCACGGA
dme-miR-184	5’ TGGACGGAGAACTGATAAGGGC
gga-miR-30a-3p	5’ CTTTCAGTCGGATGTTTGCAGC
hsa-let-7b	5’ TGAGGTAGTAGGTTGTGTGGTT
hsa-miR-1825	5’ TCCAGTGCCCTCCTCTCC
hsa-miR-423-5p	5’ TGAGGGGCAGAGAGCGAGACTTT
Control primer pairs	
X. laevis ODC	F: 5’ ACAAAGAAACCCAAACCAGA R: 5’ CAAACAACATCCAGTCTCCAA
X. tropicalis ODC	F: 5’ GGCCACACTGGCAACTCATGC R: 5’ CGGTGTGCGCTCAGTTCTGGT
X. tropicalis U6	F: 5’ ATGGCCCCTGCGCAAGGATG R: miRNA qPCR reverse primer
Human GAPDH	F: 5’ CTTCAACAGCGACACCCACT R: 5’ GTGGTCCAGGGGTCTTACTC

2.4.3 MicroRNA *in situ* hybridisation

2.4.3.1 DIG end-labelling of LNA probes

LNA oligos complementary to the mature microRNA sequence were 3’ end-labelled with digoxigenin-ddUTP using the 3’ end labelling kit (Roche). The 20 μl DIG end-labelling reaction was applied to a sephadex G25 microspin column (Amersham) to purify the probe from unincorporated DIG-ddUTP. Successful DIG end-labelling was confirmed by applying 1 μl spots of the reaction to a nylon membrane, UV cross-linking and Western blotting for DIG. The membrane was blocked for 30 minutes in

blocking reagent and incubated overnight at 4 °C in anti-DIG antibody conjugated to alkaline phosphatase (Roche) diluted 1/1000 in blocking reagent (see section 2.3.5.1). The membrane was washed in 1 x MAB with 0.1 % (v/v) Tween (see section 2.3.5.1), followed by washing in alkaline phosphatase buffer. The anti-DIG blot was developed by applying 1 ml of CDSP reagent to the membrane (Roche) and incubation at room temperature for 5 minutes to allow the luminescent reaction to occur before exposing the blot to x-ray film to obtain the signal. If end-labelling was successful, the probe was diluted to a 20 x stock (200 nM) in miRNA in situ hybridisation buffer (2.4.3.2) and stored at -20 °C.

DIG end-labelling reaction

100 pmol LNA oligo (Exiqon)

4 µl 5 x reaction buffer

4 µl $CoCl_2$ solution (25 mM)

1 µl DIG-ddUTP (1 mM)

1 µl Terminal transferase (20 units/µl)

Up to 20 µl with RNAse-free water

2.4.3.2 In situ hybridisation of LNA probes to miRNA

Xenopus embryos for microRNA *in situ* hybridisation were fixed in 4% formaldehyde in PBSA and stored in 100% methanol at -20°C until use. All washes were at room temperature unless otherwise stated. Firstly, embryos were rehydrated by successive 5 minute incubations in 75% methanol in PBST, 50% methanol in PBST, 25% methanol in PBST and finally three 5-minute incubations in 100% PBST. Embryos were treated with proteinase K (100 µl of a 10 mg/ml stock in 100 ml PBST) for 5 minutes, and then refixed for 20 minutes in 4% formaldehyde in PBS and washed in PBST three times for 5 minutes. Embryos were then washed in water for 5 minutes. Endogenous alkaline phosphatase activity was then blocked by incubation in triethanolamine/acetic

anhydride solution followed by a 5-minute water wash and three 5 minute washes in PBST.

The embryos were then transferred to hybridisation buffer and prehybridised for 2 hours at 62 °C. Next, the embryos were transferred to a 10 nM solution of antisense DIG-labelled LNA probe in hybridisation buffer and left to hybridise overnight at 62 °C. Post-hybridisation washed were carried out at 62 °C. Embryos were incubated for 10 minutes in HM- buffer, then 75 % HM-/ 25 % 2 x SSCT, then 50 % HM-/ 50 % 2 x SSCT, then 25 % HM-/ 75 % 2 x SSCT followed by 2 x SSCT. Embryos were then washed twice for 30 minutes in 0.2 x SSCT and transferred to PBST by three 5-minute washes.

The embryos were then blocked for 1 hour in blocking solution and incubated overnight at 4 °C in anti-DIG antibody conjugated to alkaline phosphatase diluted 1/2000 in blocking solution. Post-antibody washes in TBST-levamisole were as follows: six 1-hour washes at room temperature followed by overnight at 4 °C. The embryos were then washed twice for 10 minutes in alkaline phosphatase buffer and staining was developed by transferring the embryos to a 24-well plate and adding 1 ml of BM purple reagent (Roche). Staining was developed overnight at room temperature. Embryos were then bleached by adding 1 ml of bleaching solution to the well and incubating in bright light for 10 minutes to reveal the BM purple staining. Embryos were washed in PBST to remove the bleach and fixed in 4 % paraformaldehyde in PBS overnight at 4 °C.

2.4.3.3 MicroRNA in situ hybridisation solutions

PBST

1 x PBS

0.1 % (v/v) Tween-20 (Sigma)

Triethanolamine/acetic anhydride solution (100 ml)

1.49 g triethanolamine (Sigma)

Up to 100 ml with RNAse-free water

Then add 2.55 ml acetic anhydride (98 % w/v Sigma)

Hybridisation buffer

50 % (v/v) deionised formamide

5 x SSC

0.1 % (w/v) Tween

9.2 mM citric acid (Fisher)

50 μg/ml heparin (Sigma)

500 μg/ml yeast RNA (torula, Sigma)

Made up in RNAse-free water

HM- buffer

50 % (v/v) deionised formamide

5 x SSC

0.1 % (w/v) Tween-20 (Sigma)

9.2 mM citric acid (Fisher)

Made up in RNAse-free water

2 x SSCT (100 ml)

10 ml 20 x SSC (Gibco)

100 µl Tween-20 (Sigma)

Up to 100 ml with RNAse-free water

0.2 x SSCT

1 ml 20 x SSC (Gibco)

100 µl Tween-20 (Sigma)

Up to 100 ml with RNAse-free water

Blocking solution (100 ml)

20 ml 10 % (v/v) foetal calf serum

0.2 g BSA (Sigma)

Up to 100 ml in RNAse-free water

TBST-levamisole (500 ml)

50 ml 10 x TBS

0.5 ml Tween-20 (Sigma)

0.241 g levamisole (Sigma)

Up to 500 ml with RNAse-free water

10 x TBS

100 mM Tris pH 7.4

1.5 M NaCl

Made up in water, autoclaved and stored at room temperature

Alkaline phosphatase buffer

100 mM Tris-HCl pH 9.5

50 mM $MgCl_2$

100 mM NaCl

0.1 % (v/v) Tween-20 (Sigma)

Made up in RNAse-free water

Bleaching solution (100 ml)

60 ml RNAse-free water

32.5 ml hydrogen peroxide (30 % v/v Sigma)

5 ml deionised formamide (Fisher)

2.5 ml 20 x SSC (Gibco)

Up to 100 ml with RNAse-free water

2.4.4 Protein Analysis

2.4.4.1 List of antibodies

Antibody	Species and dilution used
α-XCR1 (homemade)	Rabbit, 1/5000
α-Phospho-Smad2 (Millipore AB3849)	Rabbit, 1/1000
α-Smad2/3 (BD transduction labs 610843)	Mouse, 1/1000
α-E-cadherin (BD transduction labs 610181)	Mouse, 1/2000
α-GFP (Cancer Research UK Cell Services)	Mouse, 1/1000
α-MCM-6 (Santa Cruz SC9843)	Goat, 1/2000
α-Tubulin (Abcam 6160)	Rat, 1/2000

2.4.4.2 Purification of α-XCR1 antibody from rabbit serum

Serum from a rabbit immunised with an XCR1 peptide (sequence CKSTKTLPFLGITDGKKL) was affinity-purified using a sulfolink affinity column (Pierce) with the peptide coupled to the beads via the sulfhydryl group of the N-terminal

cysteine residue of the peptide. Following equilibration of the affinity column, the diluted rabbit serum was applied to the column. Fractions were collected from the column and those containing protein were pooled and dialysed overnight against 50 % glycerol/PBS. 8 ml of purified antibody was aliquoted and tested for XCR1 specificity by Western blot. A dilution of 1/5000 was sufficient to detect endogenous XCR1.

2.4.4.3 Xenopus embryo protein extracts

10 µl of lysis buffer per embryo was added to an eppendorf tube containing snap-frozen *Xenopus* embryos. The embryos were lysed whilst thawing by repeated pipetting. The lysed embryos were centrifuged at 12,500 rpm at 4 °C to clarify the protein extract. 10 µl of extract (equivalent to protein extract from one embryo) was added to 4 x loading buffer (see section 2.4.4.6) and denatured by heating at 95 °C for 5 minutes. Protein samples were stored at -20 °C before use. Protein samples intended for XCR1 Western blotting were treated with 1 µl of PNGase F (500 units/µl) to remove N-linked sugars from proteins by incubation at 37 °C for 1 hour before protein quantification and denaturation.

Xenopus embryo lysis buffer

100 µl 1 M Tris pH 8.0

20 µl 0.5 M EDTA pH 8.0

250 µl 10 % (v/v) NP40

100 µl 1.25 M sodium-β-glycerophosphate

500 µl 1 M sodium fluoride

5 µl 20 µM Calyculin A (Cell signalling)

250 µl 200 mM sodium pyrophosphate

15 µl protease inhibitor cocktail (complete mini tablet, Roche)

Up to 5 ml with water

Prepared immediately before use

2.4.4.4 Protein quantification

Measurement of protein concentration before the addition of 4 x loading was required for dissected embryo protein samples to ensure even loading. Protein concentration was determined by preparing a standard curve of BSA (0-10 μg). 200 μl of Biorad protein assay reagent was added to the standards and the absorbance of the standards at 595 nm was measured using a Spectramax Plus spectrophotometer. The absorbance of 1 μl of the unknown protein sample at 595 nm after the addition of Biorad protein assay reagent was measured and compared to the BSA standard curve, allowing the protein concentration of the unknowns to be calculated.

2.4.4.5 Western blot analysis

Proteins were separated according to their size and mobility by SDS-polyacrylamide gel electrophoresis. Resolving gel mix of the appropriate percentage (20 % gels for GFP and XCR1 and 10 % gels for phospho-Smad2) was poured between two sealed glass plates and allowed to polymerise. The resolving gel was overlaid with water to ensure a level gel edge formed. 6 ml of stacking gel was then poured on top of the resolving gel and a comb was inserted between the glass plates. After the gel had set, the comb was removed and the gel was placed into electrophoresis apparatus. 1 x running buffer was added. Denatured protein samples in 1 x loading buffer were loaded onto the gel with a protein molecular weight marker. The gel was run at 230 V and 65 mA for 1.5 hours (phospho-Smad2), 2 hours (GFP) or 2 hours 30 minutes (XCR1).

After gel electrophoresis, the gel was opened and the stacking gel cut away. The gel was equilibrated in 1 x transfer buffer whilst the transfer apparatus was assembled. 3 pieces of Whatman paper soaked with 1 x transfer buffer were placed onto a semi-dry transfer unit, followed by a PVDF membrane (Millipore) pre-activated in methanol and equilibrated in transfer buffer. Next, the gel was placed on top of the membrane and 3 more pieces of saturated Whatman paper were placed on top of the gel. The gel sandwich was rolled to remove any air bubbles. The proteins were transferred to the membrane at 230 V, 180 mA for 1 hour. Successful transfer was confirmed by Ponceau S staining of the membrane and the positions of molecular weight protein markers were marked onto the membrane. The membrane was destained in water and washed in PBST for 5 minutes. The membrane was then blocked by rocking incubation in 5 %

milk-PBST for 30 minutes at room temperature. Next, the blocking solution was poured off and the primary antibody for the protein of interest diluted in 5 % milk-PBST was added to the membrane and incubated overnight at 4 °C on a rocker.

The membrane was then washed four times for 15 minutes in PBST and incubated for 1 hour with secondary antibody conjugated to horseradish peroxidase (DakoCytomation) diluted in 5 % (w/v) milk-PBST. The membrane was then washed four times for 15 minutes in PBST and finally washed in PBS for 5 minutes. The protein signal was then detected by applying ECL detection reagent (Amersham) to the membrane for 1 minute. The luminescent signal was detected by x-ray film or by a LAS-4000 ImageQuant machine (GE healthcare).

2.4.4.6 Western blot buffers and reagents

4 x loading buffer

4 ml 20 % (w/v) SDS

3.2 ml 1 M Tris-HCl (pH 6.8)

4 ml glycerol (Fisher)

0.1 ml β-mercaptoethanol

0.025 % (w/v) bromophenol blue

Protein molecular weight marker

A molecular weight marker was loaded alongside *Xenopus* embryo extracts containing myosin (212 kDa), β-galactosidase (116 kDa), glycogen phosphorylase (98 kDa), BSA (67 kDa), catalase (58 kDa), glutamate dehydrogenase (55 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa) and RNAse A (13 kDa) at 50 μg/ml .

10% resolving gel mix for Phospho-Smad2 gels

7.3 ml water

3.75 ml 40 % acrylamide mix

3.8 ml 1.5 M Tris pH 8.8

75 µl 20 % (w/v) SDS

40 µl 20 % (w/v) APS

30 µl TEMED

20 % resolving gel mix

7.5 ml 40 % (w/v) acrylamide

0.5 ml 2% (w/v) bis acrylamide

3.75 ml 1.5 M Tris-HCl (pH 8.8)

3.25 ml water

40 µl 20 % (w/v) APS

30 µl TEMED

Stacking gel mix

0.625 ml 40 % (w/v) acrylamide

0.333 ml 2% (w/v) bis acrylamide

0.625 ml 1.0 M Tris-HCl (pH 6.8)

3.417 ml water

15 µl 20 % (w/v) APS

10 µl TEMED

1 x Running buffer

384 mM Glycine

50 mM Tris-HCl

0.1 % (w/v) SDS

Made up in water

10 x Transfer buffer

1.5 M glycine

200 mM Tris-HCl

0.1 % (w/v) SDS

Made up in water

20 % (v/v) methanol is included in 1 x transfer buffer

Ponceau S

2 % (w/v) Ponceau S
30 % (v/v) trichloroacetic acid

5% Milk-PBST

5 g powdered milk (Marvel)

100 ml PBST

PBST

1 x PBS

0.1 % (v/v) Tween-20 (Sigma)

2.5 Microscopy

A Leica MZFLIII dissecting scope fitted with a Leica DFC420 C camera was used to photograph *Xenopus* embryos for phenotypic analysis or record *in situ* hybridisation staining patterns. For confocal microscopy analysis, fixed stage 10 embryos were placed on 1 % agarose dishes and bisected through dorsal-ventral axis by cutting

through the dorsal lip of the blastopore. Embryos were mounted on glass slides with a chamber made by cutting a window in 4 layers of black electrical tape. Molwiol mounting medium was used (Calbiochem). Embryos were imaged using an LSM-510 laser scanning microscope (Zeiss) fitted with a 10 x objective water immersion lens.

2.6 Small RNA sequencing Illumina library preparation

2.6.1 Obtaining *Xenopus tropicalis* embryo RNA

Xenopus tropicalis embryos were obtained by *in vitro* fertilisation as described in section 2.2.7. Stage 8, stage 10 and stage 18 whole embryos were snap-frozen on dry ice and stored at -80 °C. Embryos were dissected into animal and vegetal halves at stage 10 as described in 2.2.8.3 and snap-frozen on dry ice and stored at -80 °C. Total RNA was then extracted using Trizol reagent as described in section 2.3.1 and the quality of the RNA was confirmed by gel electrophoresis (see section 2.1.8.1). The RNA was stored in an isopropanol slurry at -80 °C prior to Illumina small RNA library preparation. I constructed the small RNA libraries at the Miska laboratory in the Gurdon Institute, Cambridge UK under the supervision of Javier Armisen.

2.6.2 Purification of 18-30 nt small RNAs

18-30 nucleotide small RNAs were purified from total RNA. The starting quantity of total RNA for each small RNA library was 148 μg for the stage 8, stage 10 and stage 18 libraries, 48 μg for the stage 10 animal library and 96 μg for the stage 10 vegetal library. A 15 % 8 M urea gel was poured in a 1 mm Criterion cassette (Biorad) and the gel was pre-run using 0.5 x TBE running buffer at 200 V for 1.5 hours in a mini protein gel II gel apparatus (Biorad). The total RNA slurries were divided into 50 μg samples, precipitated on dry ice for 10 minutes and centrifuged at 13,000 rpm at 4 °C to obtain RNA pellets. The pellets were resuspended in 15 μl of RNAse-free water and 15 μl of 2 x gel loading buffer II (Ambion) and denatured by heating to 80 °C for 5 minutes.

The samples were then loaded onto the gel with an 18-26 nucleotide RNA marker (Sigma). An empty lane was left between the ladder and samples and between samples for different libraries. The gel was run at 150 V until the bromophenol blue reached the bottom of the gel (about 2 hours). The ladder lane was then cut out and stained with

ethidium bromide for 5 minutes and the gel slices corresponding to 18-30 nt for each RNA sample were excised and transferred to an autoclaved non-stick eppendorf tube. 500 µl of 0.3 M sodium acetate pH 5.2 and 0.1 % (w/v) SDS solution was added and the RNA was eluted overnight by rotating the tube at 4 °C. The eluate was transferred to a new eppendorf tube and extracted using phenol-chloroform. 500 µl of phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 13,000 rpm at 4 °C for 5 minutes. The upper aqueous layer was precipitated with an equal volume of isopropanol and 1 µl of glycogen (20 mg/ml, Roche) on dry ice for 45 minutes. The samples were then centrifuged at 13,000 rpm at 4 °C for 30 minutes to obtain an RNA pellet. The pellet was washed with 80 % ethanol and dried in a vacuum centrifuge for 5 minutes on low heat. RNA samples for each library (stage 8, stage 10, stage 18, stage 10 animal and stage 10 vegetal) were recombined by resuspending sequentially in 5.7 µl of RNAse-free water.

15 % 8 M urea gel mix

9 ml Sequagel concentrate (National diagnostics)

4.5 ml Sequagel diluent

1.5 ml 10 x TBE

120 µl 10 % (w/v) APS

6 µl TEMED

2.6.3 5’ adapter ligation and purification

Next, a 26 nt RNA adapter (see section 2.6.10) was ligated onto the 5’ end of the purified small RNAs. Firstly, the purified small RNA was denatured by heating at 90 °C for 30 seconds and snap cooling on ice. The 5’ adapter ligation reaction was assembled and incubated at 37 °C for 2 hours. The reaction was stopped by the addition of 10 µl of 2 x gel loading buffer (Ambion) and denatured by heating at 80 °C for 5 minutes prior to loading on a 15 % 8M urea gel. The gel was pre-run in 0.5 x TBE for 30 minutes at 200 V. The ligated samples were loaded with a gap between samples from different libraries and 37-46 nt and 18-26 nt RNA markers (Sigma) were loaded. The gel was run

at 150 V until the bromophenol blue reached the end of the gel (about 2 hours). The marker lane was cut away and stained in ethidium bromide solution. The stained marker was replaced next to the gel and the gel slice corresponding to 42-56 nt was excised and transferred to an autoclaved non-stick eppendorf tube.

500 µl of 0.3 M sodium acetate pH 5.2 and 0.1 % (w/v) SDS solution was added and the RNA was eluted overnight by rotating the tube at 4 °C. The eluate was transferred to a new eppendorf tube and extracted using phenol-chloroform. 500 µl of phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 13,000 rpm at 4 °C for 5 minutes. The upper aqueous layer was precipitated with an equal volume of isopropanol and 1 µl of glycogen (20 mg/ml, Roche) on dry ice for 45 minutes. The samples were then centrifuged at 13,000 rpm at 4 °C for 30 minutes to obtain an RNA pellet. The pellet was washed with 80 % ethanol and dried in a vacuum centrifuge for 5 minutes on low heat. Ligated RNA samples were resuspended in 6.4 µl of RNAse-free water.

5' adapter ligation reaction

5.7 µl purified small RNA

1.3 µl 5' RNA adapter (5 µM)

1 µl 10 x RNA ligation buffer (Promega)

1 µl T4 RNA ligase (10 units/µl, Promega)

1 µl RNAsin (Promega)

2.6.4 3' adapter ligation and purification

Next, a 23 nt 3' RNA adapter (see section 2.6.10) was ligated onto the purified small RNA 5' ligation product. The purified 5' ligation product RNA was denatured by heating at 90 °C for 30 seconds and snap cooling on ice. The 3' adapter ligation reaction was assembled and incubated at 37 °C for 2 hours. The reaction was stopped by the addition of 10 µl of 2 x gel loading buffer (Ambion) and denatured by heating at 80 °C

for 5 minutes prior to loading on a 10 % 8M urea gel. The gel was pre-run in 0.5 x TBE for 30 minutes at 200 V. The ligated samples were loaded with a gap between samples from different libraries and 52-66 nt RNA markers (Sigma) were loaded. The gel was run at 150 V until the bromophenol blue reached the end of the gel (about 2 hours). The marker lane was cut away and stained in ethidium bromide solution. The stained marker was replaced next to the gel and the gel slice corresponding to 62-79 nt was excised and transferred to an autoclaved non-stick eppendorf tube.

500 µl of 0.3 M sodium acetate pH 5.2 and 0.1 % (w/v) SDS solution was added and the RNA was eluted overnight by rotating the tube at 4 °C. The eluate was transferred to a new eppendorf tube and extracted using phenol-chloroform. 500 µl of phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 13,000 rpm at 4 °C for 5 minutes. The upper aqueous layer was precipitated with an equal volume of isopropanol and 1 µl of glycogen (20 mg/ml, Roche) on dry ice for 45 minutes. The samples were then centrifuged at 13,000 rpm at 4 °C for 30 minutes to obtain an RNA pellet. The pellet was washed with 80 % ethanol and dried in a vacuum centrifuge for 5 minutes on low heat. Ligated RNA samples were resuspended in 4.5 µl of RNAse-free water.

10 % 8 M urea gel mix

6 ml Sequagel concentrate (National Diagnostics)

7.5 ml Sequagel diluent

1.5 ml 10 x TBE

120 µl 10 % (w/v) APS

6 µl TEMED

3' adapter ligation reaction

6.4 μl purified 5' ligation product

0.6 μl 3' RNA adapter (10 μM)

1 μl 10 x RNA ligation buffer (Promega)

1 μl T4 RNA ligase (10 units/μl, Promega)

1 μl RNAsin (Promega)

2.6.5 Reverse transcription of ligated small RNAs

The ligated RNA was then used as the template for cDNA synthesis by reverse transcription using the SuperScript II reverse transcription kit (Invitrogen). The reverse transcription reaction was assembled and incubated as described below.

Reverse transcription reaction

4.5 μl purified ligated small RNA

0.5 μl RT primer (100 μM)

65 °C for 10 minutes

Then add:

2 μl 5 x first strand buffer (Invitrogen)

0.5 μl 12.5 mM dNTP mix

1 μl 100 mM DTT

0.5 μl RNAsin (Promega)

48 °C for 3 minutes

Then add 1 μl of SuperScript II RT (200 units/μl, Invitrogen)

Incubate at 42 °C for 1 hour

2.6.6 Large-scale PCR for Illumina sequencing

The cDNA from the reverse transcription reaction was used as the template in a large-scale PCR reaction using PCR primers 1 and 2 (see section 2.6.10). The PCR product was then precipitated with 20 μl of 3 M sodium acetate (pH 5.2) and 500 μl of 100 % ethanol on dry ice for 30 minutes. The samples were centrifuged at 13,000 rpm for 20 minutes at 4 °C to obtain a DNA pellet. This was washed in 80 % ethanol, dried and resuspended in 20 μl of water and 5 μl of 5 x TBE-DNA loading buffer was added.

Large-scale PCR reaction

4 μl cDNA

40 μl 5 x Phusion High Fidelity polymerase buffer (NEB)

4 μl dNTP mix (10 mM)

10 μl PCR primer 1 (10 μM)

10 μl PCR primer 2 (10 μM)

2 μl Phusion DNA polymerase (2000 units/ml)

130 μl RNAse-free water

PCR program

98 °C 30 seconds

98 °C 10 seconds
58 °C 30 seconds
72 °C 20 seconds
} x 20 cycles

72 °C 5 minutes

4 °C ∞

2.6.7 Final gel purification for Illumina sequencing

A 6 % native gel was poured. The PCR samples were loaded onto the gel with each PCR sample divided into two lanes and 5 μl of 100 bp hyperladder V (Sigma) was loaded. The gel was run at 100 V in 0.5 x TBE until the bromophenol blue reached the bottom of the gel. The gel was then stained in 0.5 x TBE with ethidium bromide

solution for 5 minutes. The DNA bands were visualised on a UV trans-illuminator and the band at approximately 100 nt was excised, taking care to avoid the adapter-dimer band at around 70 nt. The 100 nt band appears smeary due to the different sizes of the small RNAs in the library. 400 µl of NEB buffer 2 was added to the excised gel slice and the DNA was eluted by rotation overnight at 4 °C. The eluate was transferred to a new eppendorf tube containing 1 µl of glycogen (Roche), 40 µl of 3 M sodium acetate pH 5.2 and 1 ml of 100 % ethanol and precipitated on dry ice for 20 minutes, then centrifuged at 13,000 rpm at 4 °C to obtain the DNA pellet. The pellet was washed with 70 % ethanol, dried, and resuspended in 15 µl of sterile resuspension buffer.

6 % native gel mix

1.5 ml 10 x TBE

4.5 ml 40 % acrylamide Accugel (19: 1 acrylamide: bis-acrylamide, National diagnostics)

24 ml water

300 µl 10 % (w/v) APS

10 µl TEMED

Resuspension buffer

10 mM Tris-HCl pH 8.5

Autoclaved and stored at room temperature.

2.6.8 Quantification of small RNA PCR products

The concentration of the resuspended PCR products was determined by Nanodrop quantification and adjusted to 10 nM using resuspension buffer. The libraries were then submitted for Illumina sequencing at the Cancer Research UK Cambridge Research Institute.

2.6.9 Illumina sequencing

The small RNA libraries were sequenced using the small RNA sequencing primer (section 2.6.10) by single-end small RNA sequencing using the Illumina platform on a Genome Analyzer IIx. The sequencing run takes 4 days and generates FastQ sequence files for downstream analysis. The small RNA libraries were sequenced a second time approximately 1 year later for a technical repeat on an Illumina Genome Analyzer IIx at the Cancer Research UK London Research Institute.

2.6.10 Illumina small RNA library oligonucleotide sequences

5' RNA Adapter (26 nt)

5' GUUCAGAGUUCUACAGUCCGACGAUC

3' RNA Adapter (23 nt)

5' P-UCGUAUGCCGUCUUCUGCUUGUidT

RT primer

5' CAAGCAGAAGACGGCATACGA

PCR Primer 1

5' CAAGCAGAAGACGGCATACGA

PCR Primer 2

5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA

Small RNA Sequencing Primer

5' CGACAGGTTCAGAGTTCTACAGTCCGACGATC

2.7 mRNA sequencing

2.7.1 Obtaining *X. tropicalis* total RNA for mRNA-seq

I obtained total RNA from stage 8, stage 10 and stage 18 *X. tropicalis* whole embryos obtained by *in vitro* fertilisation (section 2.2.7) and Trizol RNA extraction (section 2.3.1).

2.7.2 Illumina mRNA library preparation

mRNA libraries were prepared from *X. tropicalis* total RNA by the Cancer Research UK Advanced sequencing facility at the Cancer Research UK London Research Institute. 2.5 μg of total RNA was used as the input for mRNA library preparation. Briefly, mRNA was purified from total RNA using magnetic beads with poly(T) oligo attached. The mRNA was then fragmented using positively charged ionic species (Mg^{2+} and K^{+}) and heating to 94 °C for 5 minutes in Illumina 5 x fragmentation buffer (Tris-acetate, magnesium acetate and potassium acetate). The fragmented mRNA was precipitated, resuspended in water and used as the template for first strand cDNA synthesis using SuperScript II reverse transcriptase (Invitrogen). Next, the mRNA strand was removed and a replacement cDNA strand was synthesised to generate double-stranded cDNA. The ends of the double-stranded cDNA were repaired using T4 DNA polymerase and Klenow DNA polymerase. 3' overhangs were removed by 3' to 5' exonuclease activity, and 5' overhangs were filled in by polymerase activity to convert overhangs into blunt ends.

Next, a single A base was added to the 3' end of the blunt cDNA fragments using Klenow. This is to allow subsequent ligation of adapters with a single T overhang at the 5' end. The Illumina mRNA-sequencing adapters were then ligated on to the cDNA fragments using T4 DNA ligase. The products of the ligation reaction were gel purified by electrophoresis on a 2 % agarose gel and excision of a 200 bp band. The cDNA was purified from the gel slice and amplified by PCR using Phusion DNA polymerase. The PCR products were purified using the QIAquick PCR purification kit (Qiagen) and eluted in 30 μl of elution buffer provided with the kit. The concentration of the mRNA library was measured and integrity of the 200 bp band confirmed by agarose gel electrophoresis. The mRNA libraries were adjusted to a suitable concentration for Illumina sequencing, and sequenced on Illumina Genome Analyzer IIx at the Cancer Research UK London Research Institute. 72 bp of single-end mRNA sequence was obtained and deposited in FastQ files for downstream analysis.

2.8 Bioinformatic analysis

I designed the bioinformatic analysis of the small RNA sequencing and mRNA sequencing, which carried out by Stuart Horswell in the Cancer Research UK Bioinformatics and Biostatistics group. Details of bioinformatic analyses are referred to in the appropriate sections and figures within Chapters 4 and 5. Sequence files from the small RNA-seq and mRNA-seq were viewed in the Integrative Genomics Viewer browser (Robinson et al., 2011).

Chapter 3. Uncovering the mechanism of spatial regulation of *Xenopus* Cripto-1

3.1 Introduction

The aim of the work presented in this chapter was to determine the mechanism of spatial regulation of *Xenopus laevis* Cripto-1 protein, XCR1. XCR1 is a coreceptor for Nodal ligands. Therefore spatial regulation of XCR1 is critical for spatial control of Nodal signalling and thus patterning of the early *Xenopus* embryo. Firstly, I take an unbiased approach, by mapping the regulatory region of the *XCR1* gene and then go on to examine the potential regulators, such as microRNAs, that mediate spatial regulation of XCR1.

3.2 Results

3.2.1 XCR1 is subject to spatial regulation in early development

As discussed in section 1.2, XCR1 is subject to spatial regulation at the beginning of gastrulation, stage 10 (Dorey and Hill, 2006). *XCR1* mRNA is expressed in all areas of the embryo at stage 10, but the protein is absent from the vegetal pole (prospective endoderm), indicating that XCR1 is subject to post-transcriptional spatial regulation. Firstly, I corroborated these results by showing that *XCR1* mRNA is found in all areas of the embryo by RNAse protection assay with an antisense probe to the coding region of *XCR1* (Figure 3.1). I showed that XCR1 protein is expressed in the animal cap and marginal zone, but not the vegetal pole of embryos at both stage 10 and stage 8 (Figure 3.2). In addition to confirming the published result, the observation that XCR1 is spatially regulated at stage 8 suggests that a maternal factor may be responsible for spatial regulation, as stage 8 is before the onset of zygotic transcription at stage 9.

3.2.2 Functional relevance of spatial regulation of XCR1

In order to assess the function of spatial regulation of XCR1 protein in the *Xenopus laevis* embryo, XCR1 protein was reintroduced into the vegetal pole by targeted injection of 100 pg of *XCR1* mRNA into the four most vegetal blastomeres at the 32-cell stage. Effective targeting of this region was demonstrated by injection of fluorescein dextran (Figure 3.3A). The *XCR1* construct contained only the coding

region and an internal Flag tag and no *XCR1* 3' or 5' UTRs. XCR1 was replaced in the vegetal pole at a level comparable to endogenous XCR1 in the animal cap-marginal zone region, as shown by Western blot for XCR1 using dissected embryo extracts (Figure 3.3B).

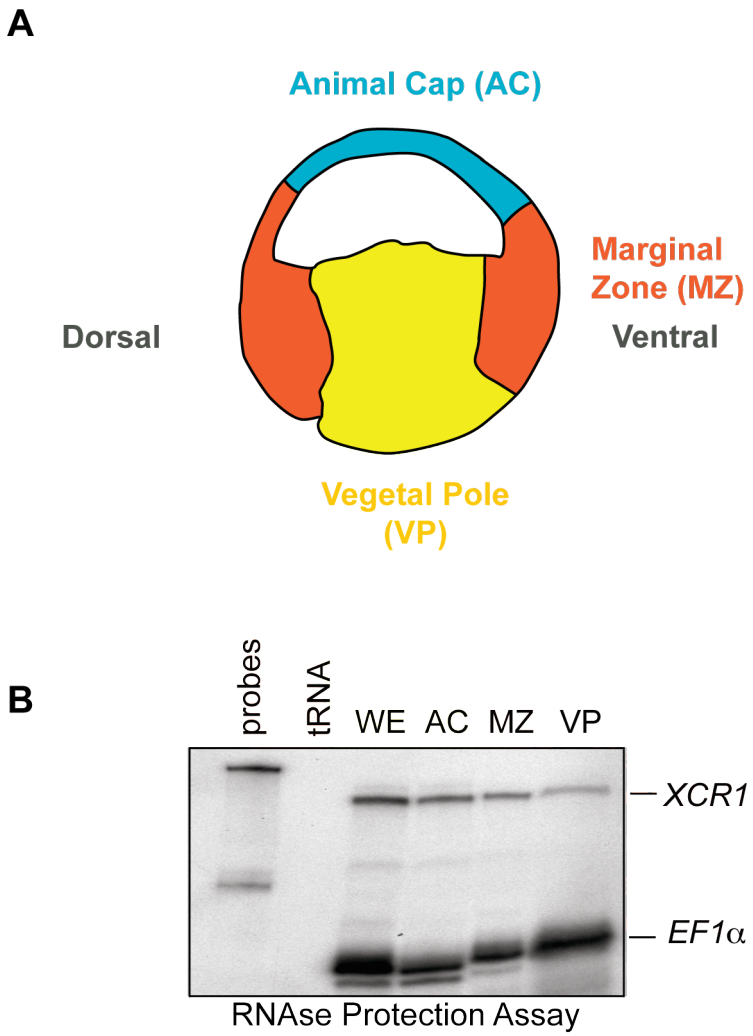


Figure 3.1 *XCR1* RNA is expressed in all areas of the *Xenopus laevis* embryo at stage 10
A. Lateral view schematic of the gastrula stage embryo at stage 10 illustrating the animal cap (AC), marginal zone (MZ) and vegetal pole (VP) at stage 10 which can be dissected. **B.** Levels of *XCR1* and *EF-1α* mRNA were detected by RNAse protection assay. *EF-1α* serves as a loading control. Antisense probes for *XCR1* and *EF-1α* are shown prior to RNAse digestion. 20 µg of *X. laevis* stage 10 total RNA from the whole embryo (WE), animal cap (AC), marginal zone (MZ) and vegetal pole (VP) was analysed by RNAse protection with the *XCR1* and *EF-1α* probes. In the tRNA lane, no *X. laevis* embryo RNA was added. The protected fragments for *XCR1* and *EF-1α* after RNAse digestion are shown.

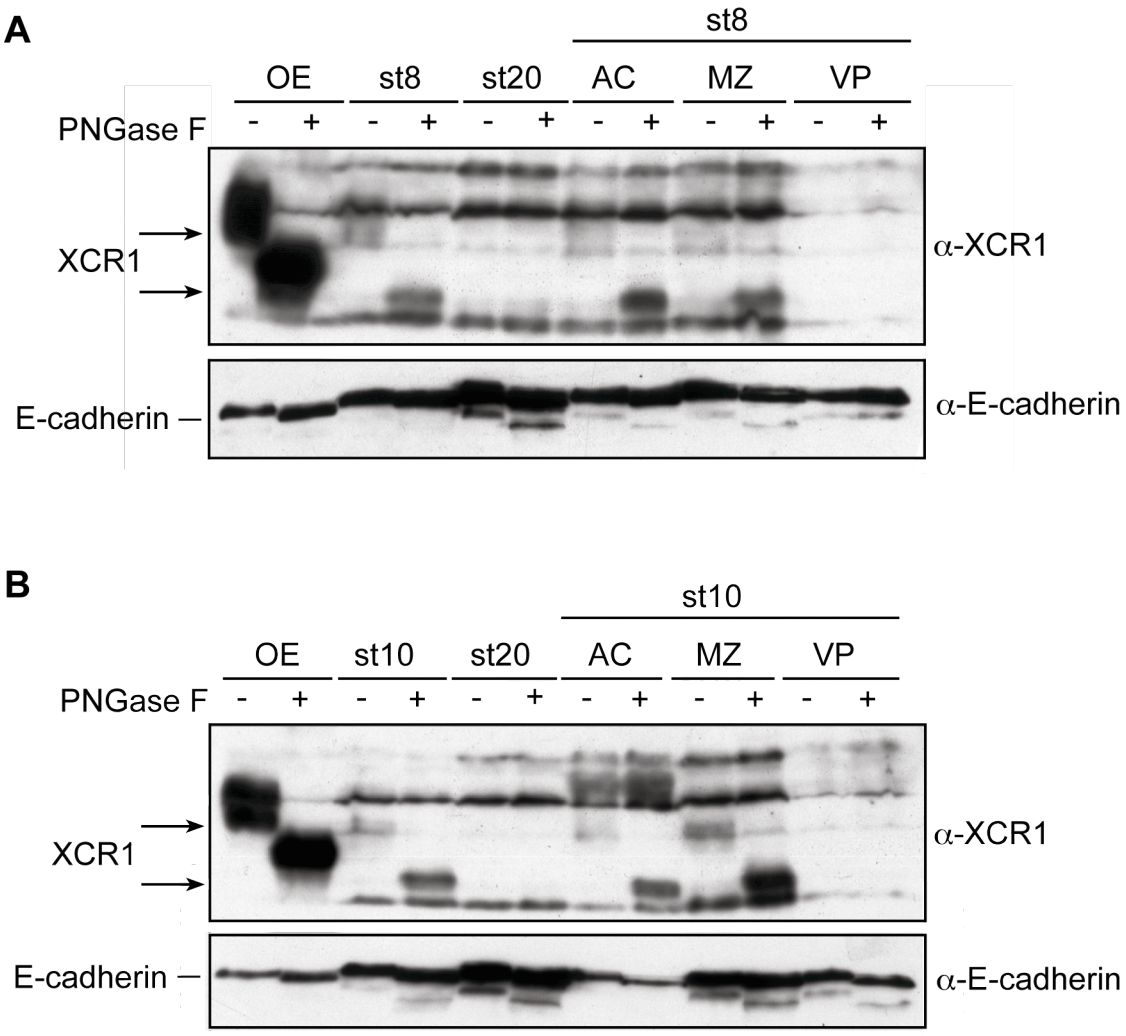


Figure 3.2 XCR1 protein is expressed in the animal cap and marginal zone, but not in the vegetal pole at stage 8 and stage 10
A. Western blot for XCR1 protein at stage 8. Whole embryo or dissected embryo protein extracts were treated with PNGase F to remove N-linked sugars from XCR1 to improve the resolution of XCR1 bands. XCR1 overexpression (OE) serves as a positive control for the blot, and endogenous levels of XCR1 at stage 8 and stage 20 (no XCR1 protein detected) are shown. Embryos were dissected into animal caps (AC), marginal zones (MZ) and vegetal poles (VP). XCR1 protein is absent from the vegetal pole at stage 8. **B.** Western blot for XCR1 protein at stage 10. Lanes as in A, except endogenous XCR1 levels at stage 10 are shown. XCR1 protein is absent from the VP at stage 10.

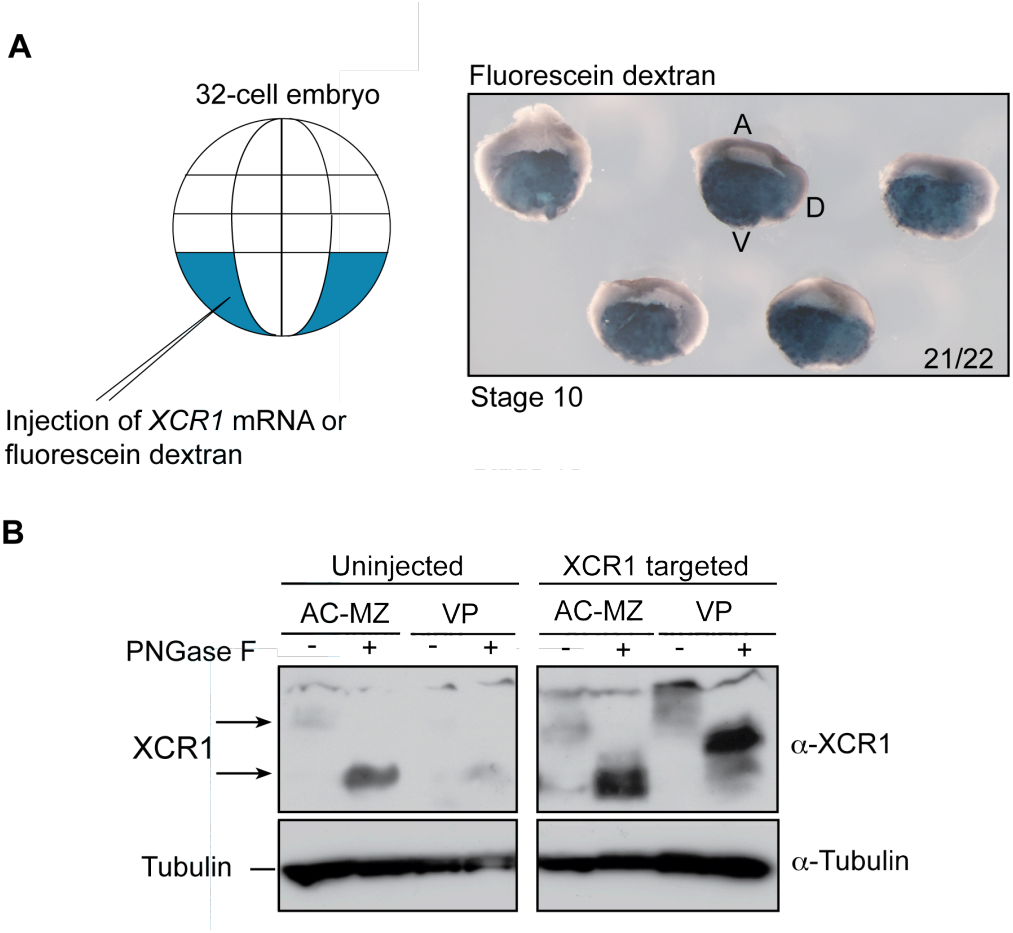


Figure 3.3 Spatial regulation of XCR1 can be ablated by vegetally targeted expression of *XCR1* mRNA

A. Schematic diagram of the 32-cell stage *Xenopus* embryo (lateral view) with the targeted blastomeres indicated in blue. The four most vegetal blastomeres are injected to target the prospective endoderm. *In situ* hybridisation for fluorescein dextran on bisected stage 10 embryos shows effective targeting of the prospective endoderm. The number of embryos out of the total analysed with the staining pattern displayed is shown. Animal (A), vegetal (V) and dorsal (D) regions are indicated. **B.** Western blot for XCR1 protein at stage 10. XCR1 is replaced at endogenous levels by targeted injection of a total dose of 100 pg of *XCR1* mRNA. Uninjected and XCR1 targeted embryos were dissected into animal cap-marginal zone and vegetal pole regions and protein extracts were untreated or treated with PNGase F. Note that the exogenous band for XCR1 in the targeted VP lane runs with lower mobility than endogenous XCR1 due to an internal Flag tag.

Embryos with ablated spatial regulation of XCR1 were allowed to develop to stage 32 and assessed for phenotypic changes compared to uninjected embryos. As shown in Figure 3.4, embryos with XCR1 vegetally targeted showed dorsalisation (11/30 embryos), defects in head development (2/30 embryos) and partial secondary axis formation (1/30 embryos). However, this phenotype was not completely penetrant, as 16/30 embryos were normal.

To look at the effect of ablating spatial regulation on Nodal signalling, embryos were dissected and extracts were prepared for Western blotting for phospho-Smad2, the downstream transducer of Nodal signalling (Schier and Shen, 2000). Embryos were dissected at stage 9, when the phospho-Smad2 signal is not yet present (Faure et al., 2000) to assess if the timing of Nodal signalling was affected by vegetal replacement of XCR1, and stage 11 when levels of phospho-Smad2 are high. As shown in Figure 3.5, vegetal replacement of XCR1 led to a detectable phospho-Smad2 signal at stage 9, in comparsison to no signal in uninjected embryos. At stage 11, the endogenous phospho-Smad2 signal is strong in uninjected embryos, and vegetal replacement of XCR1 does not markedly enhance the intensity of the signal. Ablation of XCR1 spatial regulation therefore brings forward the timing of Nodal signal transduction via phospho-Smad2.

Next, the effect of XCR1 vegetal replacement on germ layer markers was assessed by whole-mount *in situ* hybridisation at stage 10 (Figure 3.6). The molecular markers *Forkhead, Lefty, Goosecoid* and *Sox17* are Nodal target genes (Bennett et al., 2007; Dickmeis et al., 2001). *Goosecoid* is a transcriptional repressor of the mesoderm-inducing transcription factor, *Xbra*, which responds to activin signalling (Latinkic and Smith, 1999; Smith et al., 1991). Embryos with XCR1 targeted to the vegetal pole/prospective endoderm showed no change in the dorsal endoderm marker *Hhex* (Newman et al., 1997) or the pan-endodermal marker *Sox17α* (Hudson et al., 1997). Injected embryos showed strong downregulation of *Xbrachyury*, a mesodermal marker, and upregulation of the Nodal inhibitor *Lefty* (Meno et al., 1999). The mesodermal markers *Forkhead* and *Goosecoid* were moderately downregulated.

The upregulation of *Lefty* expression could be explained by the increased levels of vegetal phospho-Smad2 when XCR1 is targeted to the vegetal pole, as *Lefty* is a Nodal target gene (Saijoh et al., 2000). I hypothesise that upregulation of *Lefty* inhibits the

activation of ALK4 (through binding of Lefty to Nodal and antagonism of EGF-CFC coreceptors (Chen and Shen, 2004)), resulting in less downstream phosphorylation of R-Smads, which in turn inhibits the activation of Nodal target genes, such as the mesodermal markers, *Xbrachyury, Goosecoid* and *Forkhead*.

To determine if the phenotype of XCR1 vegetal replacement was largely due to the increase in phospho-Smad2 levels and premature onset of Smad2 phosphorylation, similar targeting experiments were performed with a constitutively-active ALK4 receptor construct, ALK4*(Armes and Smith, 1997). Ectopic expression of ALK4* elevates endogenous phospho-Smad2 levels by constitutive phosphorylation of R-Smads in the absence of ligand. Firstly, the effect of ALK4 overexpression on phospho-Smad2 levels was confirmed by injection of 200 pg of ALK4* mRNA at the 1-cell stage and Western blotting for phospho-Smad2. Ectopic expression of ALK4* caused a strong phospho-Smad2 signal at stage 9, when the endogenous phospho-Smad2 signal was not yet detectable in the whole embryo (Figure 3.7A).

25 pg of ALK4* mRNA was then targeted to the vegetal pole and embryos were dissected into animal and vegetal hemispheres at stage 9 and blotted for phospho-Smad2. Endogenous levels of Smad2 in the vegetal pole at stage 10 were also analysed to serve as a positive control for vegetal phospho-Smad2 detection. Vegetal targeting of ALK4* had similar effects on Smad2 phosphorylation to vegetal targeting of XCR1: vegetal phospho-Smad2 was strongly increased in ALK4* targeted embryos compared to uninjected embryos (Figure 3.7B). Embryos with vegetally-targeted ALK4* were allowed to develop until stage 33 and compared to uninjected embryos for phenotypic changes. As shown in Figure 3.8, vegetal targeting of ALK4* largely phenocopies vegetal targeting of XCR1. ALK4* targeted embryos displayed dorsalisation (6/31 embryos), head development defects (1/31 embryos) and partial axis duplication (1/31 embryos) (Figure 3.4), reminiscent of the phenotype of XCR1 vegetal targeting. Again, the phenotype was not fully penetrant as 16/30 ALK4* targeted embryos were normal. ALK4* targeted embryos displayed an additional phenotype of left-right axis curvature (5/31 embryos).

The striking similarity of the phenotypes of XCR1 and ALK4* vegetal targeting suggests that the functional consequences of ablating spatial regulation of XCR1 are

mediated through phospho-Smad2, namely increased phosphorylation of Smad2 and the premature onset of Nodal signal transduction at stage 9.

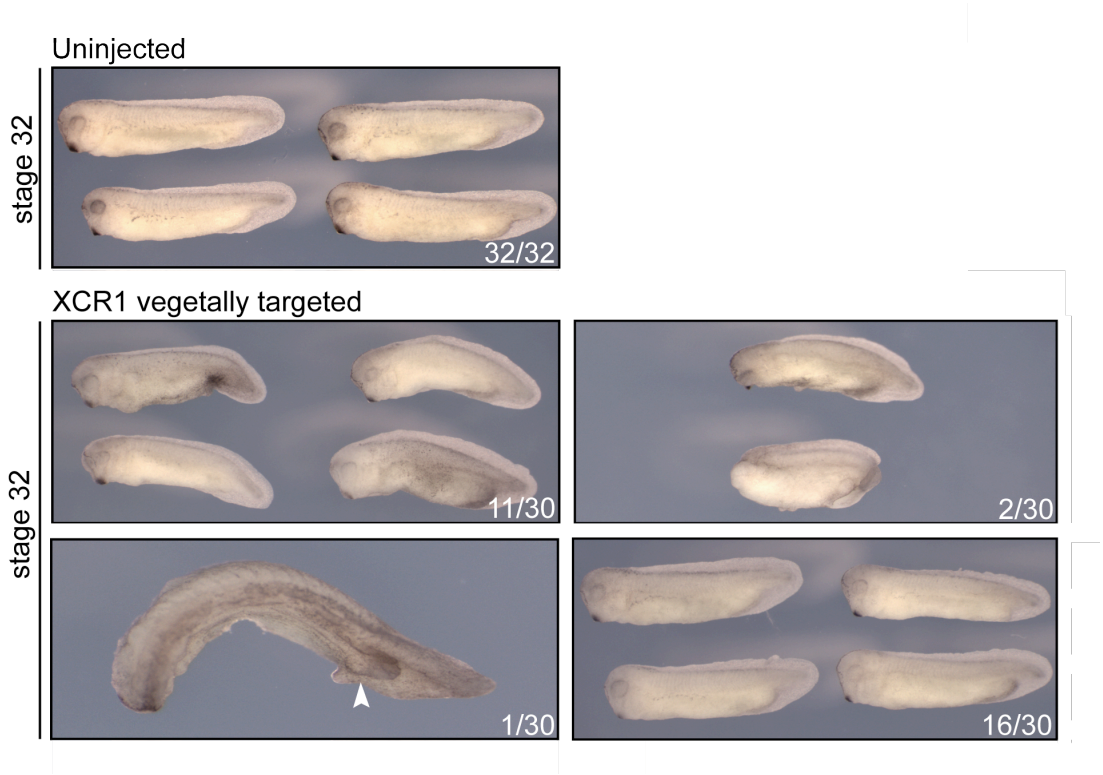


Figure 3.4 Ablating spatial regulation of XCR1 leads to dorsalisation, head development defects and secondary axis formation
Uninjected embryos and embryos with 100 pg of *XCR1* mRNA vegetally targeted at the 32-cell stage were allowed to develop until stage 32. The number of embryos with the phenotype shown out of the total number analysed is indicated. An arrowhead indicates a partial secondary axis on an XCR1 vegetally targeted embryo.

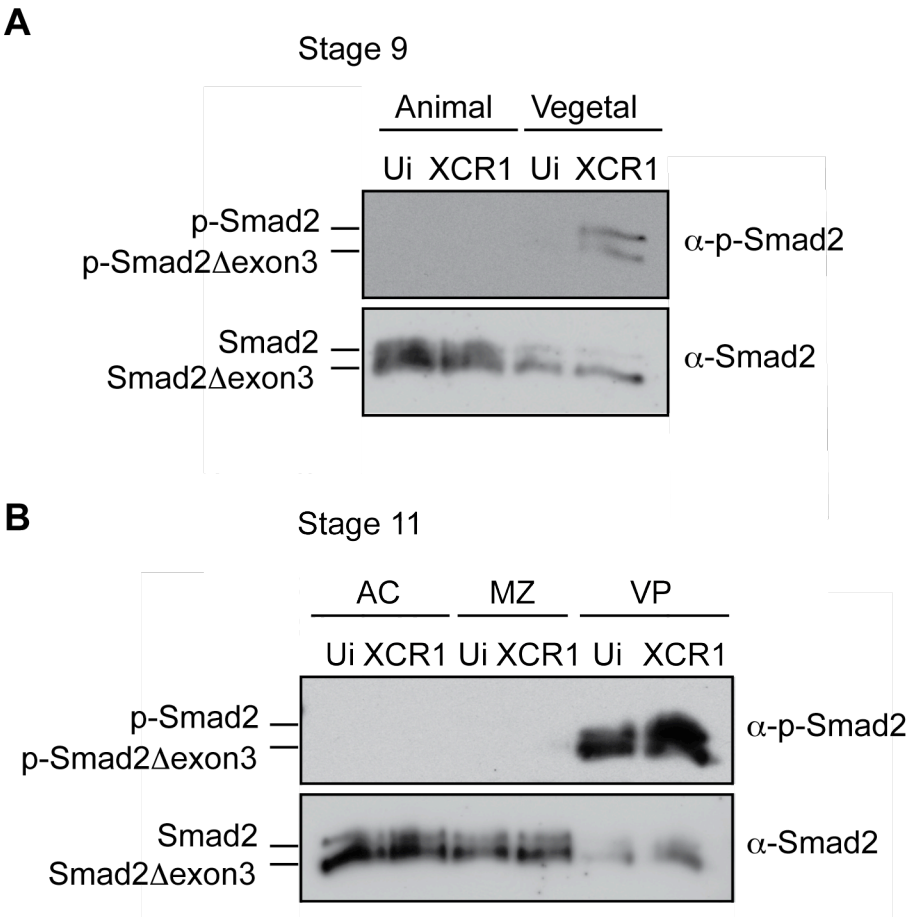


Figure 3.5 Ablation of spatial regulation of XCR1 brings forward the onset of Smad2 phosphorylation

A. Western blot for phospho-Smad2 at stage 9. Embryos were uninjected (Ui) or injected with 100 pg of *XCR1* mRNA (XCR1) in the four most vegetal blastomeres at the 32-cell stage to revert spatial regulation of XCR1 and then dissected into animal and vegetal hemispheres at stage 9. Smad2 is used as a loading control. The Smad2 isoform Smad2Δexon3 (Faure et al., 2000) is also detected and migrates below Smad2. **B.** Western blotting for phospho-Smad2 at stage 11. Embryos were uninjected (Ui) or injected with 100 pg of *XCR1* mRNA targeted to the vegetal pole at the 32-cell stage (XCR1) and dissected into animal caps (AC), marginal zones (MZ) and vegetal poles (VP) at stage 11. Smad2Δexon3 is also detected.

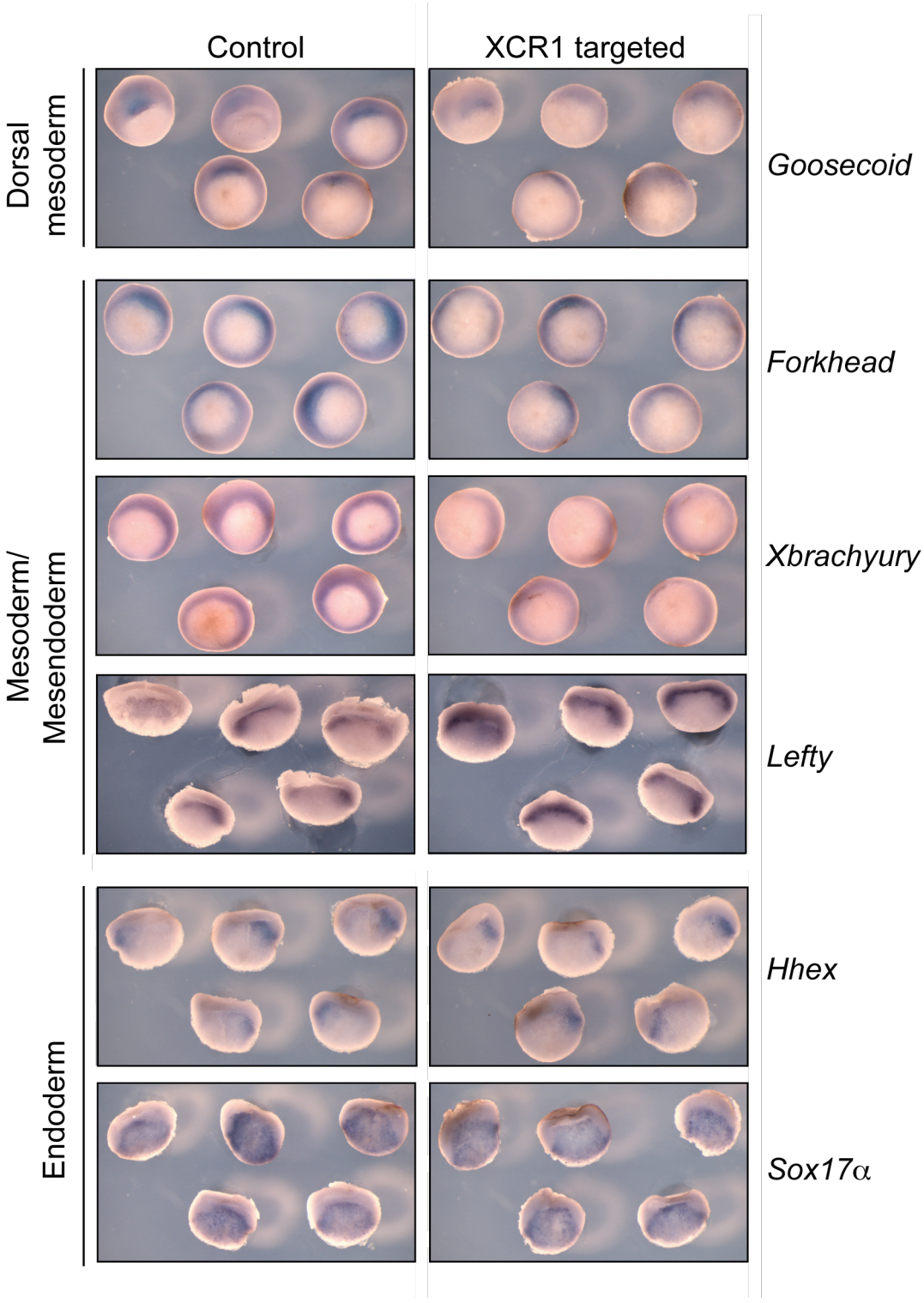


Figure 3.6 XCR1 vegetal replacement affects mesendoderm markers
Whole-mount *in situ* hybridisation for germ layer markers at stage 10. Embryos are uninjected (Control) or are subjected to targeted vegetal replacement of XCR1 at the 32-cell stage (XCR1 targeted) as in Figure 3.3. The total number of embryos analysed was 5 for each marker.

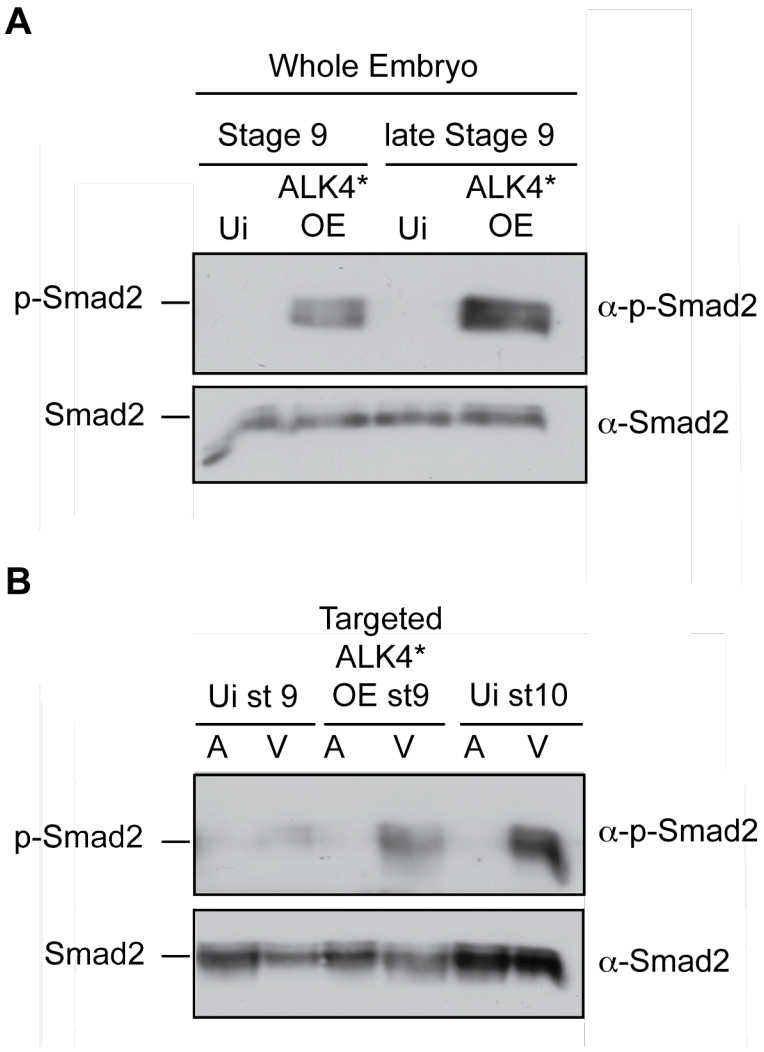


Figure 3.7 Vegetally targeted expression of constitutively active ALK4 recapitulates the effect of XCR1 vegetal replacement on Smad2 phosphorylation
A. Western blot for phospho-Smad2. 200 pg of constitutively active *ALK4* receptor mRNA was injected into *Xenopus laevis* embryos at the 1-cell stage and Smad2 phosphorylation in uninjected (Ui) and injected (ALK4* OE) embryos was analysed. Smad2 serves as a loading control. Smad2Δexon3 has not resolved on this gel. **B.** Western blot for phospho-Smad2 as in **A**. A total dose of 25pg of constitutively-active *ALK4* mRNA was targeted to the four most vegetal blastomeres at the 32-cell stage. Uninjected and injected embryos were then dissected into animal (A) and vegetal (V) hemispheres at stage 9. Stage 10 A and V lanes serve as a control for vegetal phospho-Smad2 detection.

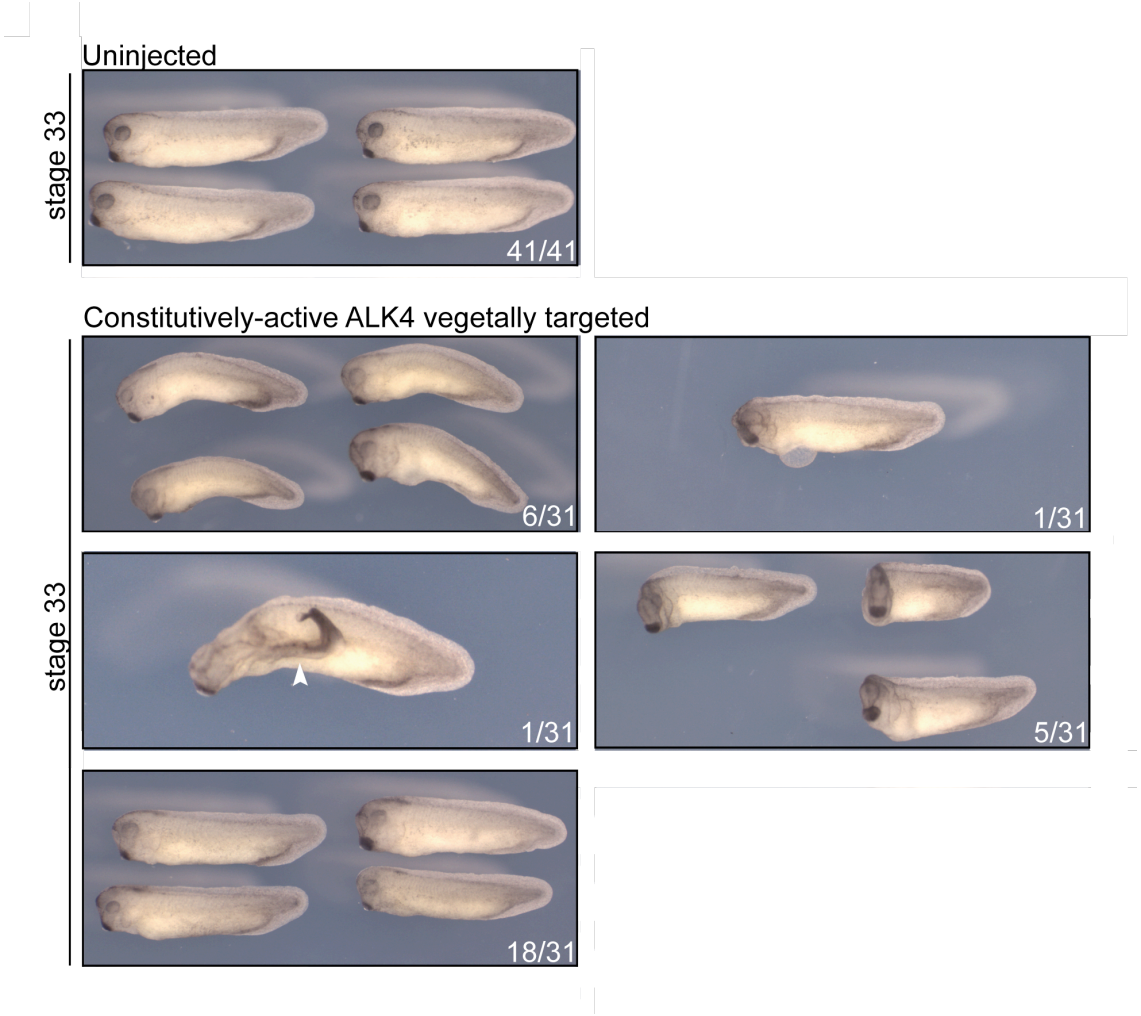


Figure 3.8 Vegetal targeting of constitutively-active ALK4 phenocopies vegetal replacement of XCR1
25 pg of constitutively-active ALK4 receptor mRNA was targeted to the four most vegetal cells at the 32-cell stage and embryos were allowed to develop to stage 33. Phenotypes of uninjected control embryos and injected embryos were scored. The number of embryos with the phenotype shown out of the total analysed is indicated. An arrowhead marks the position of a second axis on an ALK4* vegetally overexpressed embryo.

3.2.3 Spatial regulation of XCR1 is mediated by the 3'UTR of *XCR1α*

Thus far, spatial regulation of XCR1 has been shown to play a role in maintaining correct timing and magnitude of Nodal signalling pre-gastrulation. Next, I investigated the mechanism of spatial regulation of XCR1 by mapping the region of the *XCR1* gene responsible for spatial regulation. Firstly, I determined whether *XCR1* was subject to 3'UTR-mediated regulation. *Xenopus laevis* is pseudotetraploid, so there are two pseudoalleles of *XCR1*: *XCR1α*, and *XCR1β*. A sequence alignment of the *XCR1α, XCR1β*, *Xenopus tropicalis (Xtcr1)* and human *Cripto-1* 3'UTRs is shown in Figure 3.9. As described in Table 3.1, pairwise alignment scores for conservation of 3'UTR sequence between frog species are similar, whereas the *Xtcr1* 3'UTR sequence is more similar to the human 3'UTR than are the *Xenopus laevis* sequences.

To confirm that *XCR1α* and *XCR1β* were the major *XCR1* transcripts in *Xenopus laevis* and to identify any uncharacterised transcripts, Northern blot analysis on *X. laevis* and *X. tropicalis* total RNA with an antisense probe to the highly conserved *XCR1α* coding region was performed. As shown in Figure 3.10, there are three *X. laevis XCR1* transcripts, the most abundant of which is similar in size to the *X. tropicalis* transcript. From the known sizes of the *XCR1α* and *XCR1β* transcripts (1.7 kB and 1.4 kB respectively) and the similarity in size to the only *X. tropicalis* transcript (1.5kB), *XCR1α* is most likely the major *X. laevis Cripto-1* transcript. A less abundant, longer *X. laevis XCR1* transcript was also identified.

Next, the expression profiles of *XCR1α* and *XCR1β* mRNA during early *X. laevis* development were compared by RNAse protection assay (Figure 3.11). *XCR1α* mRNA was more abundant than *XCR1β* mRNA in the oocyte, and in blastula, gastrula and neurula stages of early development. This suggests that *XCR1α* and *XCR1β* are subject to differential regulation at the mRNA level, either through transcriptional regulation, or post-transcriptional regulation of mRNA stability.

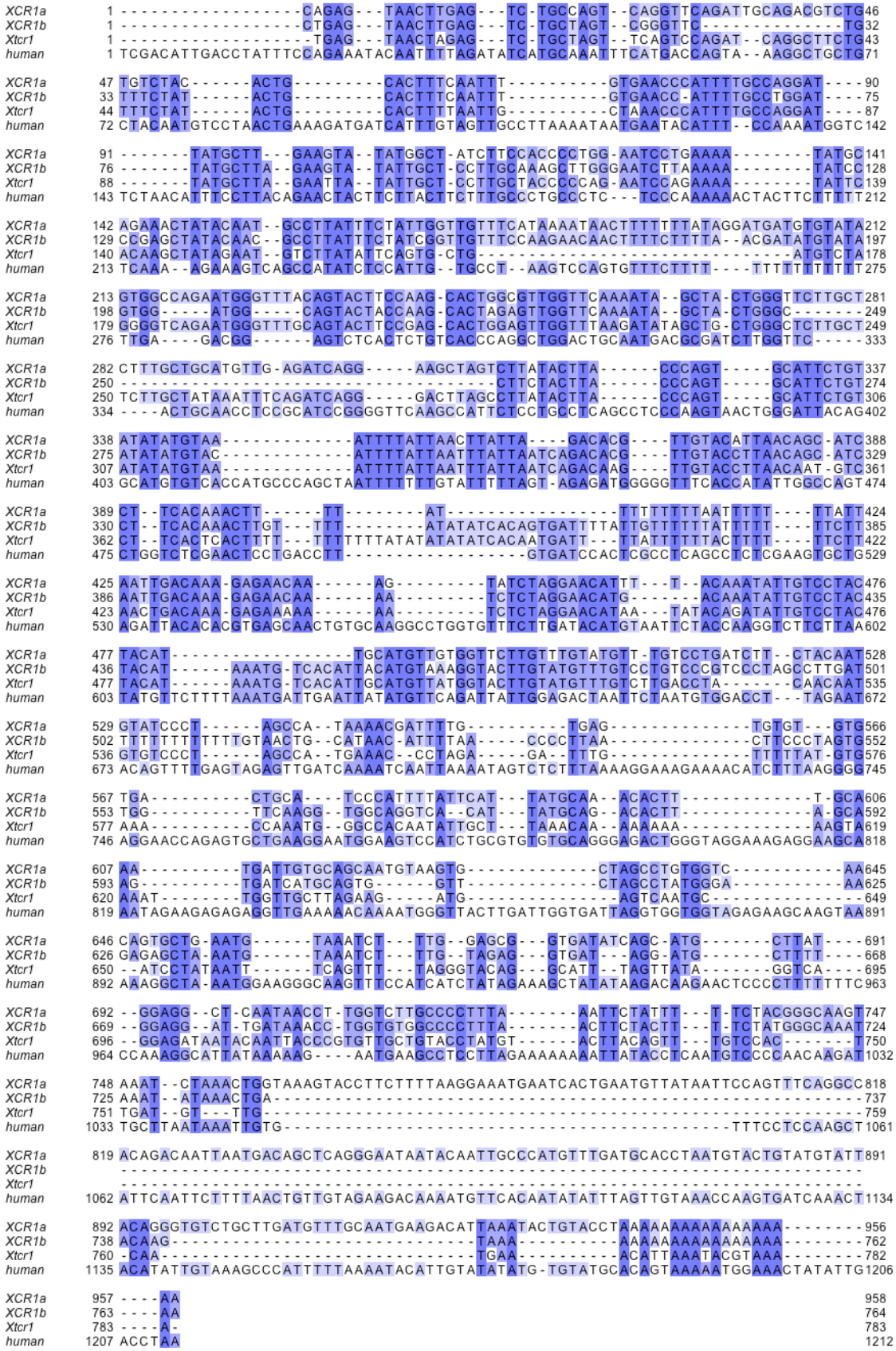


Figure 3.9 Multiple sequence alignment of *Xenopus* and human *Cripto-1* 3'UTR sequences coloured for percentage identity

3' UTR Sequences Aligned	Pairwise Alignment Score
XCR1α: XCR1β	77
XCR1α: Xtcr1	78
XCR1α: human	57
XCR1β: Xtcr1	70
XCR1β: human	59
Xtcr1: human	63

Table 3.1 *Cripto-1* 3'UTR pairwise sequence alignment scores
Cripto-1 3'UTR sequences from *Xenopus laevis* (*XCR1α and XCR1β*), *Xenopus tropicalis* (*Xtcr1*) and *Homo sapiens* (human) were aligned in Clustal W (Chenna et al., 2003) and scored for pairwise alignment. Pairwise scores are calculated from percentage identities with gap positions in the alignment excluded.

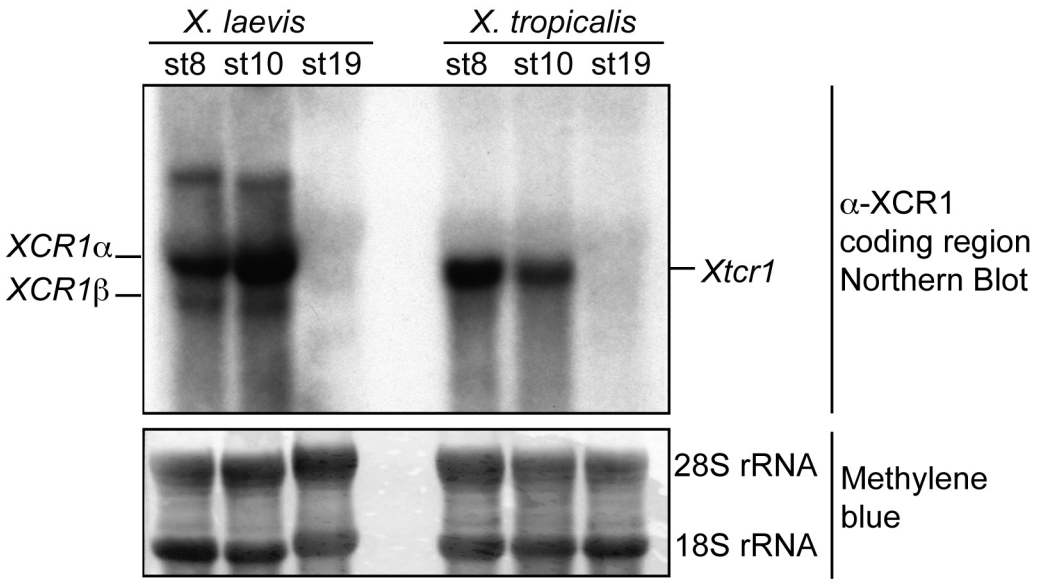


Figure 3.10 Northern blot of *Xenopus Cripto-1* transcripts
20 μg of *Xenopus* total RNA was probed with an antisense RNA probe to 600 nucleotides of the *XCR1* coding region, which is highly conserved between *XCR1α* and *XCR1β*. There is one *Cripto-1* transcript in *Xenopus tropicalis*, and three transcripts in *Xenopus laevis*, of which the major transcript is similar in size to the *X. tropicalis* transcript and may correspond to *XCR1α*. Methylene blue staining of ribosomal RNA is used as a loading control.

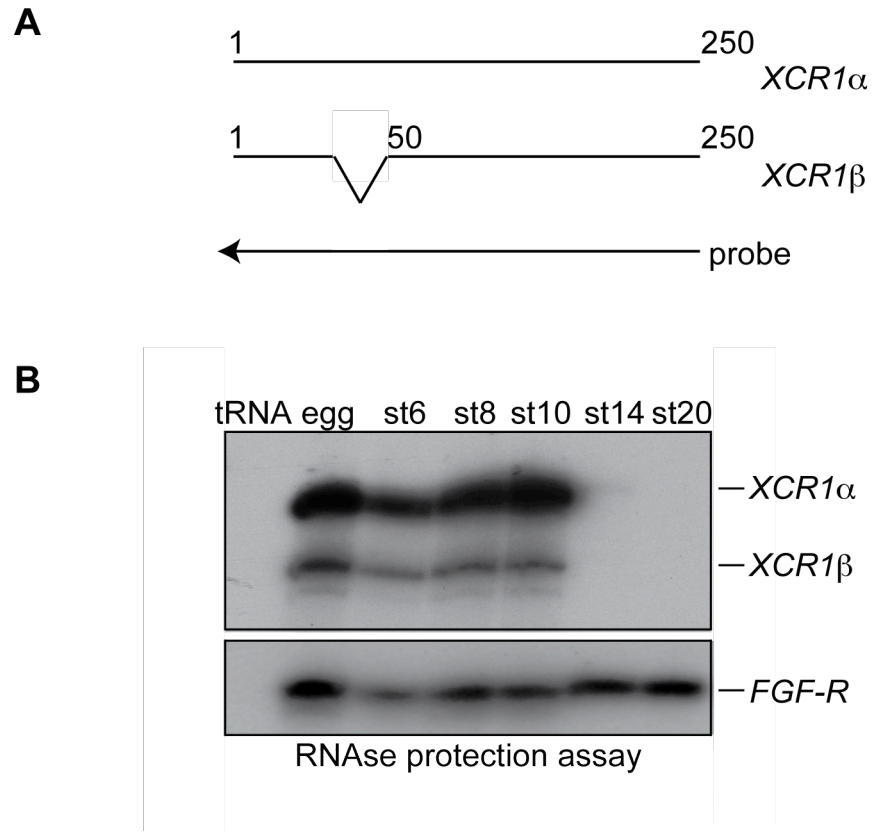


Figure 3.11 *XCR1α* mRNA is more abundant than *XCR1β* in early *Xenopus* development
A. Schematic diagram of an antisense probe to nucleotides 1-250 of the *XCR1α* coding region. The *XCR1β* sequence is highly similar to the *XCR1α* sequence, except for a 13 nt mismatch at 50 nt (indicated). The mismatch allows RNAse T1 digestion to occur, and therefore this probe yields protected fragments of 250 nt for *XCR1α* mRNA and 200 nt for *XCR1β* mRNA. **B.** 20 μg of *X. laevis* total RNA was analysed with the probe described in **A.**, to yield distinct digestion products for *XCR1α* and *XCR1β*. This allows for a quantitative comparison of *XCR1α* and *XCR1β* mRNA levels. In the tRNA lane, no *X. laevis* embryo RNA was added. *FGF receptor* mRNA levels were used as a loading control.

3.2.3.1 GFP-3'UTR reporter mRNA assay

To determine whether spatial regulation of XCR1 was mediated through the 3'UTR of *XCR1*, GFP-3'UTR reporters were cloned for the 3'UTR sequences of *XCR1α*, *XCR1β* and *Xenopus tropicalis* Cripto-1 (*Xtcr1*) (section 2.1.5). To assess the conservation of any potential spatial regulators of XCR1 identified in *Xenopus*, a GFP-3'UTR reporter for human *Cripto-1* was also cloned. Reporter plasmids were linearised downstream of the 3'UTR and synthetic mRNA was generated by *in vitro* transcription. GFP protein levels and fluorescence provide a read-out of 3'UTR-mediated regulation by endogenous *Xenopus laevis* regulators in this assay, because endogenous factors can bind the 3'UTR of the reporter and regulate GFP expression. GFP-3'UTR reporter mRNAs were coinjected with an mRNA encoding cherry-histone-H2B into *Xenopus laevis* embryos at the 4-cell stage. This assay was optimised to obtain maximum spreading of injected mRNA by injection of each blastomere in the marginal zone at the 4-cell stage. Injected embryos were allowed to develop until stage 10 (when spatial regulation of endogenous XCR1 occurs) and prepared for confocal microscopy analysis by fixation and bisection along the dorso-ventral axis (Figure 3.12). If the 3'UTR sequence downstream of GFP mediates spatial regulation, GFP fluorescence would be expected to follow the pattern of XCR1 protein localisation shown in Figure 3.2, i.e., GFP expression in the animal cap and marginal zone, but not in the vegetal pole.

Figure 3.13 shows that the 3'UTR of *XCR1α* confers spatial regulation of GFP expression, whereas the 3'UTR of *XCR1β* mediates repression of GFP expression, as fluorescence is almost undetectable. The 3'UTR of *Xtcr1* also mediates spatial regulation of GFP expression in *Xenopus laevis* embryos, which suggests that the same regulator can bind both the *X. laevis XCR1α* and *X. tropicalis Cripto-1* 3'UTRs. GFP fluorescence was observed in all areas of the embryo when the *GFP-human Cripto-1* 3'UTR reporter mRNA was injected, suggesting that spatial regulators of *Cripto-1* found in *Xenopus* are not conserved, or are not sufficient to confer spatial regulation of human *Cripto-1*. This does not, however, preclude microRNA-dependent regulation of human *Cripto-1* in mammalian systems. Spatial regulation mediated by the *XCR1α* 3'UTR is definitely a result of translational control, as injected *GFP-XCR1α 3'UTR*

mRNA is stable and present in all areas of the embryo at stage 10, as shown in Figure 3.14A.

The lack of fluorescence in *GFP-XCR1β 3'UTR* mRNA injected embryos shown in Figure 3.13, and lower endogenous *XCR1β* mRNA levels compared to *XCR1α* mRNA (Figure 3.11), suggested that *XCR1β* mRNA is less stable than *XCR1α* mRNA. To assess if the lack of *GFP-XCR1β 3'UTR* reporter signal was due to degradation of injected reporter mRNA by endogenous factors in the embryo, post-injection time courses of *GFP-XCR1α 3'UTR* mRNA and *GFP-XCR1β 3'UTR* mRNA were analysed by RNAse protection against the coding region of *GFP* (Figure 3.14B). The level of *GFP-XCR1α 3'UTR* mRNA is relatively stable until after stage 10, when the *GFP* mRNA follows the profile of endogenous XCR1 protein, which is not expressed beyond stage 12. In contrast, *GFP-XCR1β 3'UTR* mRNA is almost entirely degraded between stage 9 and stage 10. The minimal level of *GFP-XCR1β 3'UTR* mRNA at stage 10 accounts for the lack of GFP fluorescence seen in Figure 3.13, and suggests that the 3'UTR of *XCR1β* mediates instability of the endogenous *XCR1β* transcript.

In order to map the sequence in the 3'UTR of *XCR1β* that was responsible for mRNA instability, a partial GFP-3'UTR reporter clone lacking nucleotides 1-299 of the 3'UTR was made. As shown in Figure 3.15, this reporter mRNA showed restored GFP expression, indicating that nucleotides 1-299 of the *XCR1β* 3'UTR confer mRNA instability and degradation, when compared to the full length *XCR1β* 3'UTR. This result was not pursued further by finer mapping of the regulatory region within nucleotides 1-299.

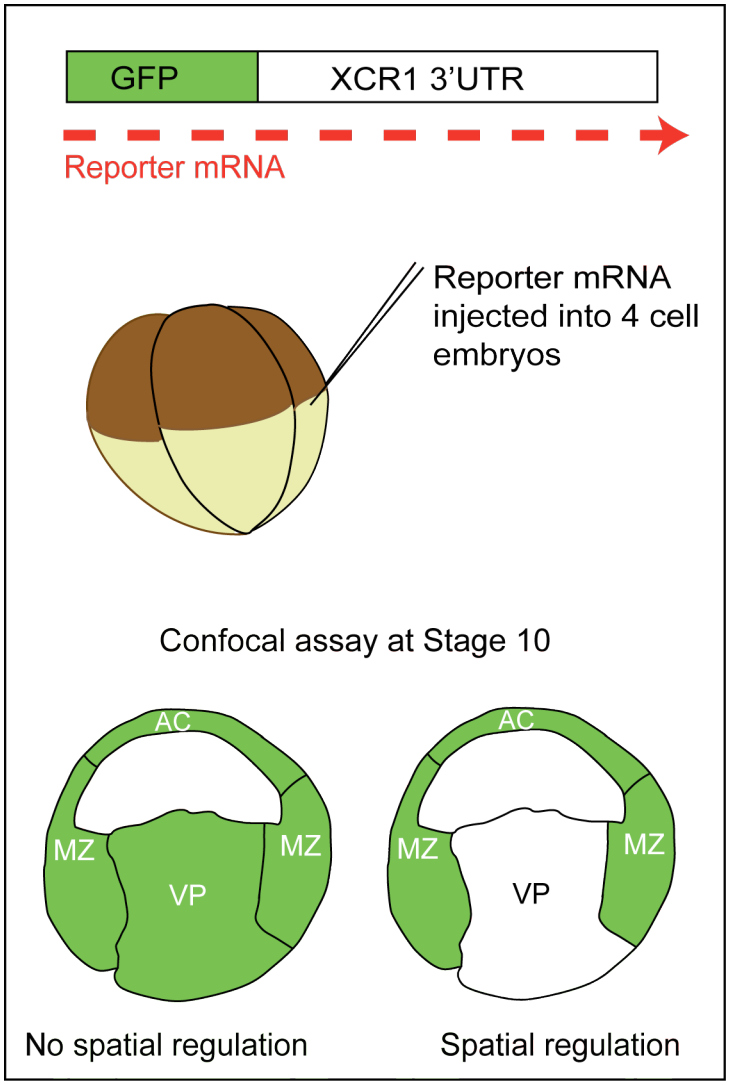


Figure 3.12 A GFP-3'UTR reporter assay for spatial regulation of XCR1
The 3'UTR of interest was cloned downstream of the coding region of *GFP* to construct a GFP-3'UTR reporter plasmid which was used as the template for *in vitro* transcription of a *GFP-3'UTR* reporter mRNA. A total dose of 800 pg of reporter mRNA was coinjected with 700 pg of *mCherry-histone-H2B* mRNA into the *Xenopus laevis* embryo at the 4-cell stage. Embryos were allowed to develop to stage 10, when they were fixed, bisected along the dorso-ventral axis and mounted for confocal microscopy analysis. GFP fluorescence provides a read-out of regulation mediated by the 3'UTR. If the 3'UTR sequence conferred spatial regulation of GFP expression, fluorescence would be expected in the animal cap (AC) and marginal zone (MZ), but not in the vegetal pole (VP), to recapitulate the localisation of endogenous XCR1 protein shown in Figure 3.2.

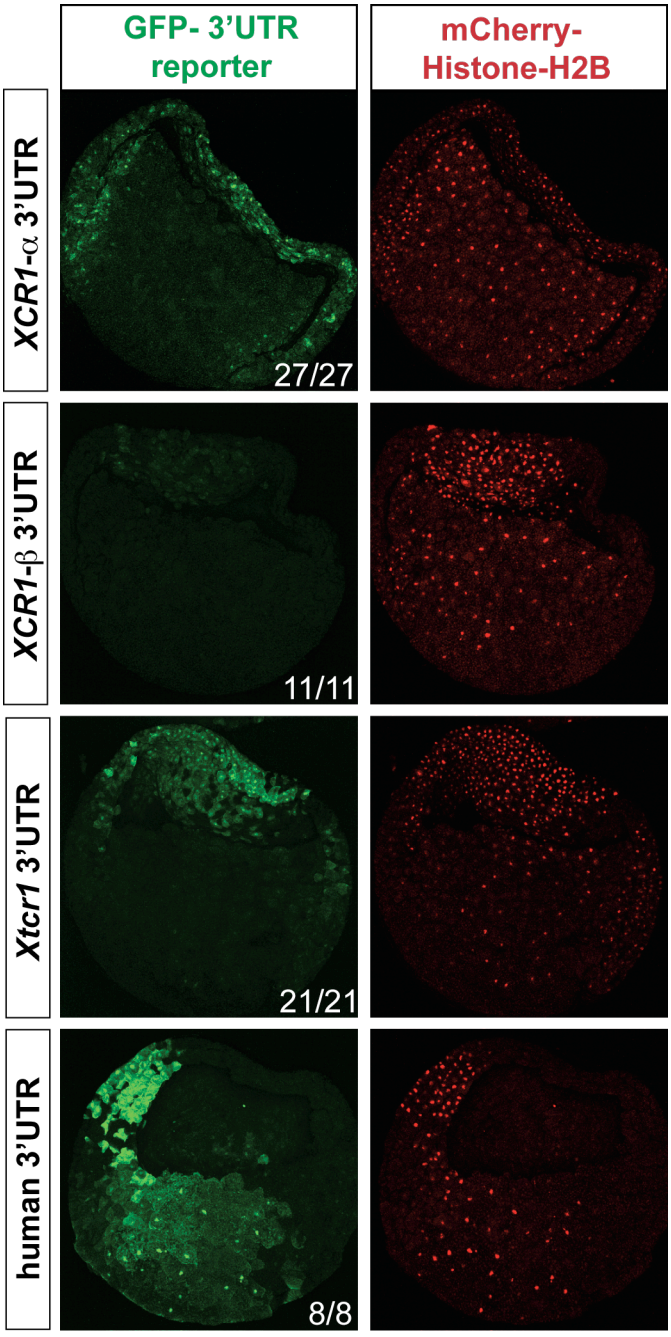


Figure 3.13 The 3' UTRs of *XCR1α* and *Xtcr1*, but not *XCR1β* and human *Cripto-1* confer spatial regulation of GFP expression
Stage 10 *X. laevis* embryos injected with *GFP-3'UTR* reporter mRNAs for the *XCR1α, XCR1β, Xtcr1* and human *Cripto-1* 3'UTRs are shown with mCherry-Histone-H2B expression as a control for spread of the injected mRNA. The total number of embryos with the GFP expression pattern shown out of those analysed is indicated.

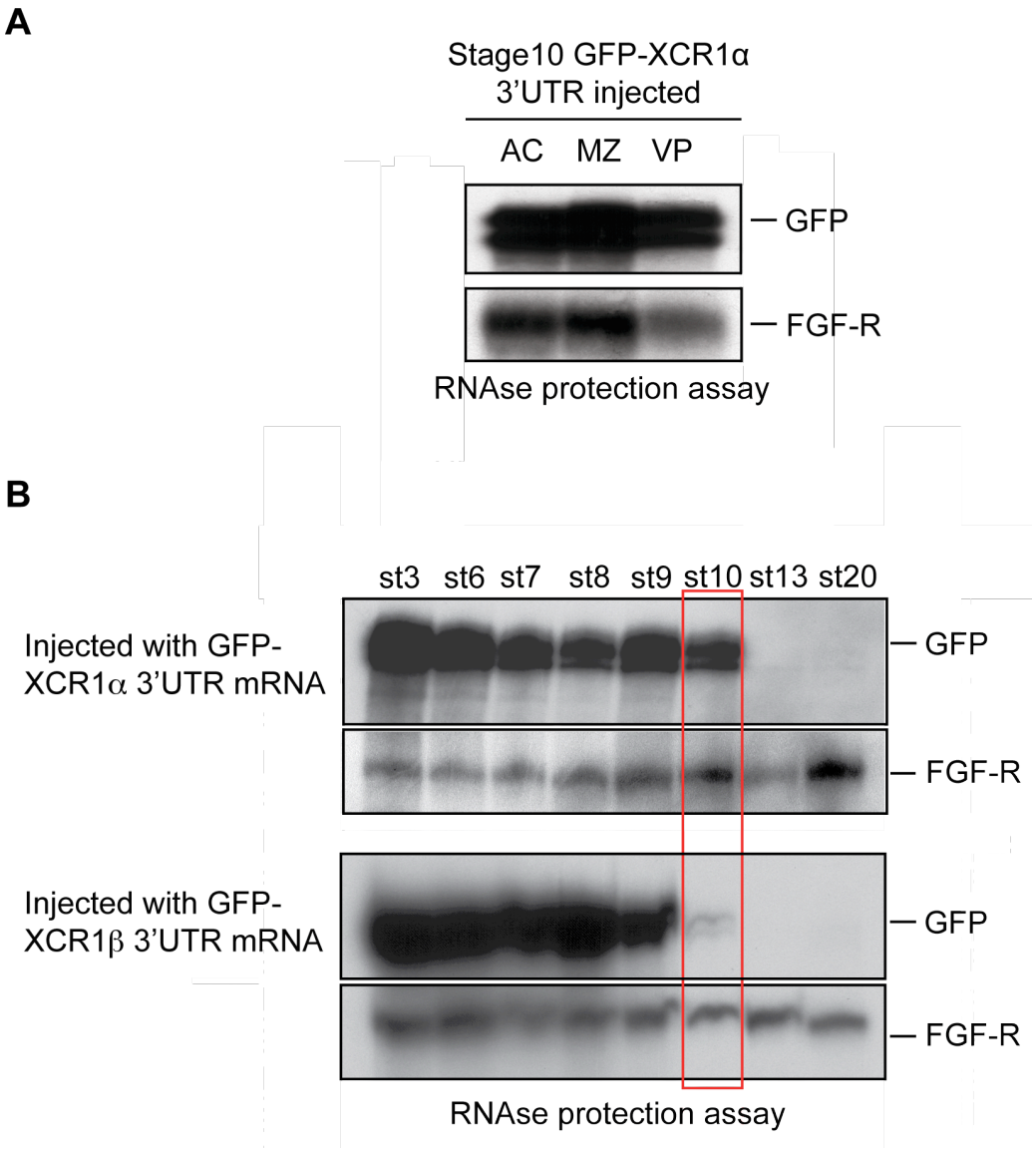


Figure 3.14 Spatial regulation mediated by the *XCR1α* 3'UTR is a translational effect, whereas the *XCR1β* 3'UTR confers mRNA instability

A. Embryos injected with *GFP-XCR1α 3'UTR* reporter mRNA at the 4-cell stage were dissected into animal caps (AC), marginal zones (MZ) and vegetal poles (VP) at stage 10 and an RNAse protection assay was performed using a *GFP* probe. *GFP* mRNA is found in all areas of the embryo, in contrast to the protein, which is only expressed in the animal cap and marginal zone (Figure 3.13). FGF receptor (FGF-R) was used as a loading control and the protected fragments are indicated. **B.** Post-injection *GFP* mRNA time course for *GFP-XCR1α 3'UTR* and *GFP-XCR1β 3'UTR* reporter mRNAs. *GFP-XCR1α 3'UTR* mRNA is present post-injection at stage 3 (4-cell stage) and is stable up until stage 10 and thereafter follows the expression profile of endogenous XCR1 protein, which is not expressed from stage 12 onwards. In contrast, *GFP-XCR1β 3'UTR* mRNA is more unstable and is degraded between stage 9 and stage 10.

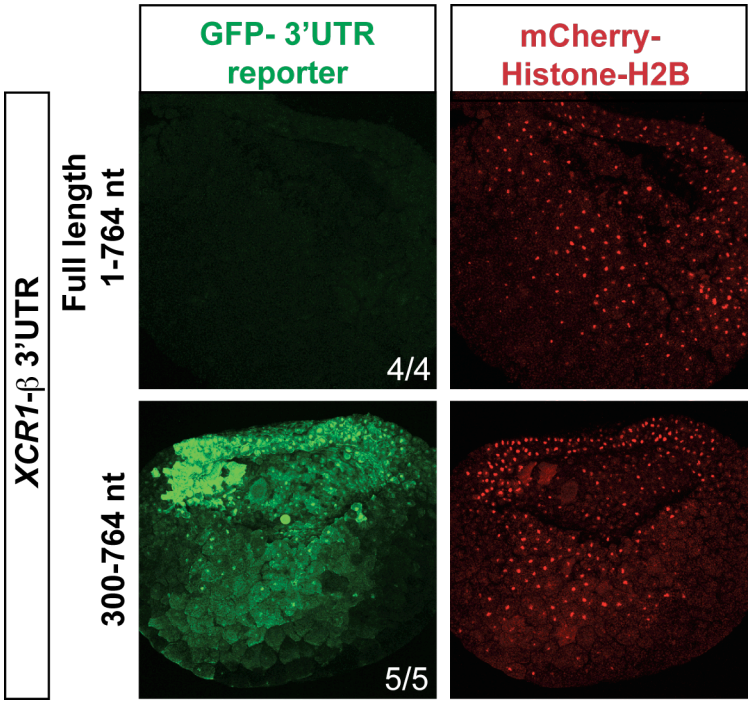


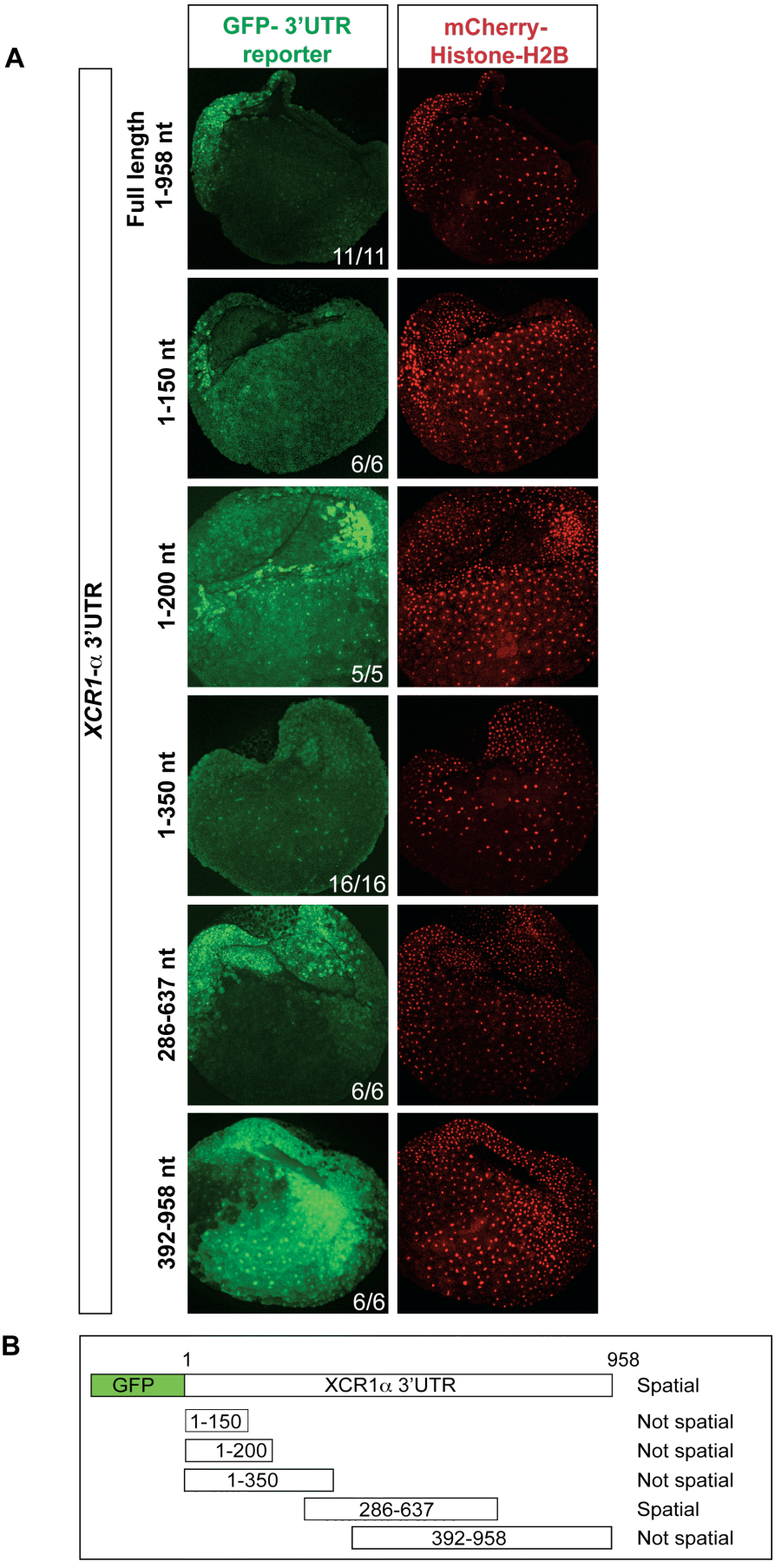
Figure 3.15 Instability of *XCR1β* mRNA is conferred by nucleotides 1-299 of the 3' UTR

Embryos were injected with *GFP-XCR1β 3'UTR* mRNA containing either the full length 3'UTR sequence, or a partial 3'UTR sequence containing nucleotides 300-764 of the 3'UTR and analysed for GFP fluorescence at stage 10. Excluding nucleotides 1-299 from the *GFP-XCR1β 3'UTR* reporter mRNA restores mRNA stability and GFP protein expression. The total number of embryos with the GFP expression pattern shown out of the total analysed is shown. mCherry-Histone-H2B expression is a control for spread of the injected mRNA.

3.2.4 Mapping the critical region of the *XCR1α* 3'UTR

The 3'UTR of *XCR1-α* has been shown to confer spatial regulation of GFP expression in a GFP-3'UTR reporter assay at stage 10 in *Xenopus laevis* embryos. This recapitulates the effect of spatial regulation of endogenous XCR1 protein at this time in development. Next, the GFP-3'UTR reporter assay was extended to map the critical region of the *XCR1-α* 3'UTR responsible for spatial regulation. An overlapping set of GFP-partial 3'UTR reporters containing nucleotides 1-150, 1-200, 1-350, 286-637 and 392-958 of the *XCR1-α* 3'UTR cloned downstream of GFP was analysed (Figure 3.16A).

As shown in Figure 3.16B, nucleotides 1-150 of the *XCR1-α* 3'UTR conferred weak expression of GFP, whereas nucleotides 1-200 conferred strong GFP expression in all areas of the embryo, i.e., non-spatial expression. Nucleotides 1-350 conferred weak, non-spatial expression of GFP. The 5' 350nt of the *XCR1-α* 3'UTR is therefore not sufficient to confer spatial regulation. A central region of the 3'UTR containing nucleotides 286-637 conferred strong, spatial regulation of GFP expression. The 3' end of the 3'UTR from nucleotides 392-958 imparted strong, non-spatial expression of GFP. The variation in GFP protein levels between the 1-150, 1-350, and 286-637 clones and the spatial regulation mediated by the *GFP-286-637 3'UTR* reporter mRNA was confirmed as an effect on translation of *GFP* mRNA, as injected *GFP* mRNA levels are stable in all areas of the embryo (Figure 3.16C).


A
GFP- 3'UTR reporter
mCherry-Histone-H2B
XCR1-α 3'UTR
Full length 1-958 nt
11/11
1-150 nt
6/6
1-200 nt
5/5
1-350 nt
16/16
286-637 nt
6/6
392-958 nt
6/6
B
1
958
GFP
XCR1α 3'UTR
Spatial
1-150
Not spatial
1-200
Not spatial
1-350
Not spatial
286-637
Spatial
392-958
Not spatial

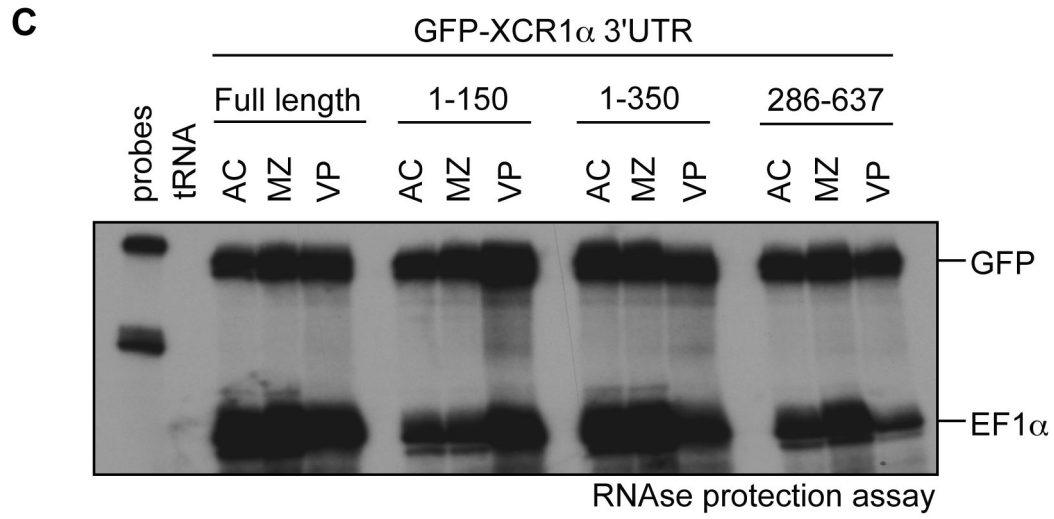


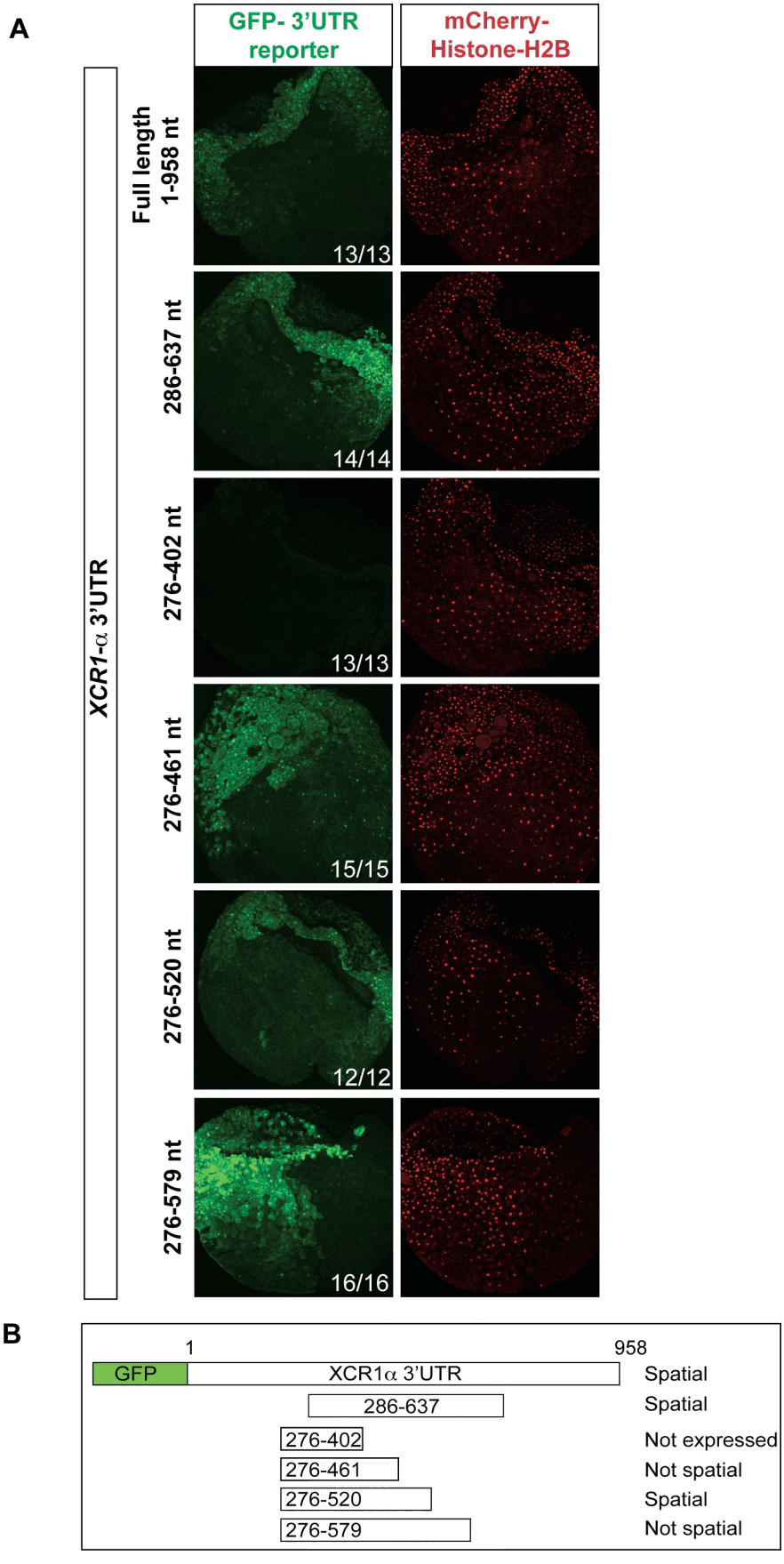
Figure 3.16 Spatial regulation of XCR1 is a translational effect mediated by nucleotides 286-637 of the *XCR1α* 3'UTR

A. *X. laevis* embryos were injected with full length or GFP-partial 3'UTR reporter mRNAs containing nucleotides 1-150, 1-200, 1-350, 286-637 and 392-958 of the *XCR1-α* 3'UTR and analysed by cofocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is shown. mCherry-Histone-H2B expression is a control for spread of the injected mRNA. **B.** Schematic representation of *GFP-XCR1-α* partial 3'UTR reporter clones and summary of results. 'Spatial' indicates that the partial 3'UTR sequence conferred spatial regulation of GFP expression by confocal microscopy. **C.** RNAse protection assay. *X. laevis* embryos were injected with *GFP-XCR1-α* partial 3'UTR reporter mRNAs and dissected into animal caps (AC), marginal zones (MZ) and vegetal poles (VP) at stage 10. Total RNA was then isolated and analysed by RNAse protection assay for *GFP* mRNA levels. *EF-1α* was used as a loading control. The protected fragments are indicated. Probes are shown prior to RNAse digestion (probes) and in the tRNA lane, no *Xenopus* embryo RNA was added.

The critical region for spatial regulation, nucleotides 286-637 of the *XCR1-α* 3'UTR, was next further divided into partial 3'UTR clones (Figure 3.17A) and analysed for spatial regulation by confocal microscopy and GFP protein levels by Western blot (Figure 3.17B,C). Whereas the full length *XCR1-α* 3'UTR and nucleotides 286-637 confer spatial regulation of GFP expression, variable GFP levels and localisation are observed with partial 3'UTR sequences. It is important to note at this point that GFP with no 3'UTR is unstable as no fluorescence is detectable.

Firstly, no GFP fluorescence was observed when nucleotides 276-402 were cloned downstream of GFP. When 60 nucleotides of downstream *XCR1-α* 3'UTR sequence are added on to this sequence in the 276-461 GFP-3'UTR clone, GFP expression is restored, but is not spatially regulated. The addition of another 60 nucleotides of *XCR1-α* 3'UTR sequence in the 276-520 3'UTR clone now confers spatial regulation of GFP expression. This suggests that nucleotides 276-520 are the minimal region sufficient to confer spatial regulation and that there is a sequence between nucleotides 461 and 520 that causes spatial regulation. A further 60 nucleotides of sequence reverts spatial regulation in the 276-579 clone, which must be counteracted by a sequence between nucleotides 579 and 637, as the 286-637 clone is spatially regulated. The effects observed with the 276-579 clone may be due to an alteration in the secondary structure of the GFP-3'UTR reporter mRNA.

In short, the minimal region for spatial regulation is 276-520, which contains a sequence conferring spatial regulation between 461-520 nucleotides. GFP protein levels in *GFP-XCR1-α partial 3'UTR* reporter mRNA injected embryos are averaged by Western blot in Figure 3.17C. This confirms the expression levels of GFP observed by confocal microscopy.


A
GFP- 3'UTR reporter
mCherry-Histone-H2B
XCR1-α 3'UTR
Full length 1-958 nt
13/13
286-637 nt
14/14
276-402 nt
13/13
276-461 nt
15/15
276-520 nt
12/12
276-579 nt
16/16
B
1
958
GFP
XCR1α 3'UTR
Spatial
286-637
Spatial
276-402
Not expressed
276-461
Not spatial
276-520
Spatial
276-579
Not spatial

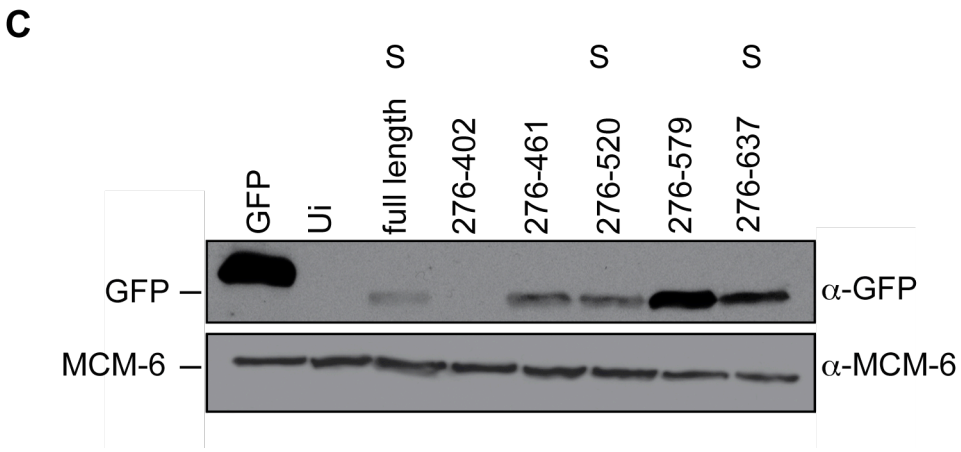


Figure 3.17 Nucleotides 276-520 of the *XCR1α* 3'UTR are sufficient to confer spatial regulation of GFP expression
A. *X. laevis* embryos were injected with full length or GFP-partial 3'UTR reporter mRNAs containing nucleotides 286-637, 276-402, 276-461, 276-520 and 276-579 of the *XCR1-α* 3'UTR and analysed by cofocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is shown. mCherry-Histone-H2B expression is a control for spread of the injected mRNA. **B.** Schematic representation of *GFP-XCR1-α* partial 3'UTR reporter clones and summary of results. 'Spatial' indicates that the partial 3'UTR sequence conferred spatial regulation of GFP expression by confocal microscopy. **C.** Western blot for GFP. Protein extract was prepared from 20 embryos injected with *GFP* mRNA, uninjected (Ui) or injected with GFP-partial 3'UTR reporter mRNAs and blotted for GFP. GFP levels confirm the expression levels observed by confocal microscopy in **A.** GFP-partial 3'UTR clones that conferred spatial regulation marked 'S'. MCM-6 is used as a loading control. Overexpressed GFP control extracts run with lower mobility than GFP-3'UTR fusions because the control GFP contains extra coding sequence.

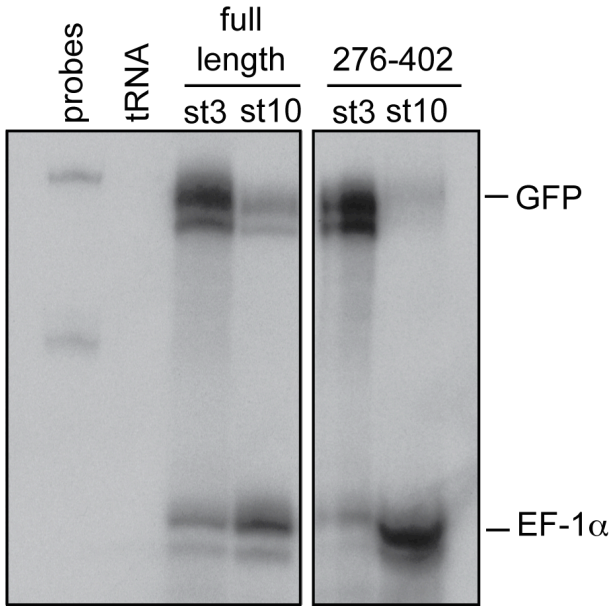


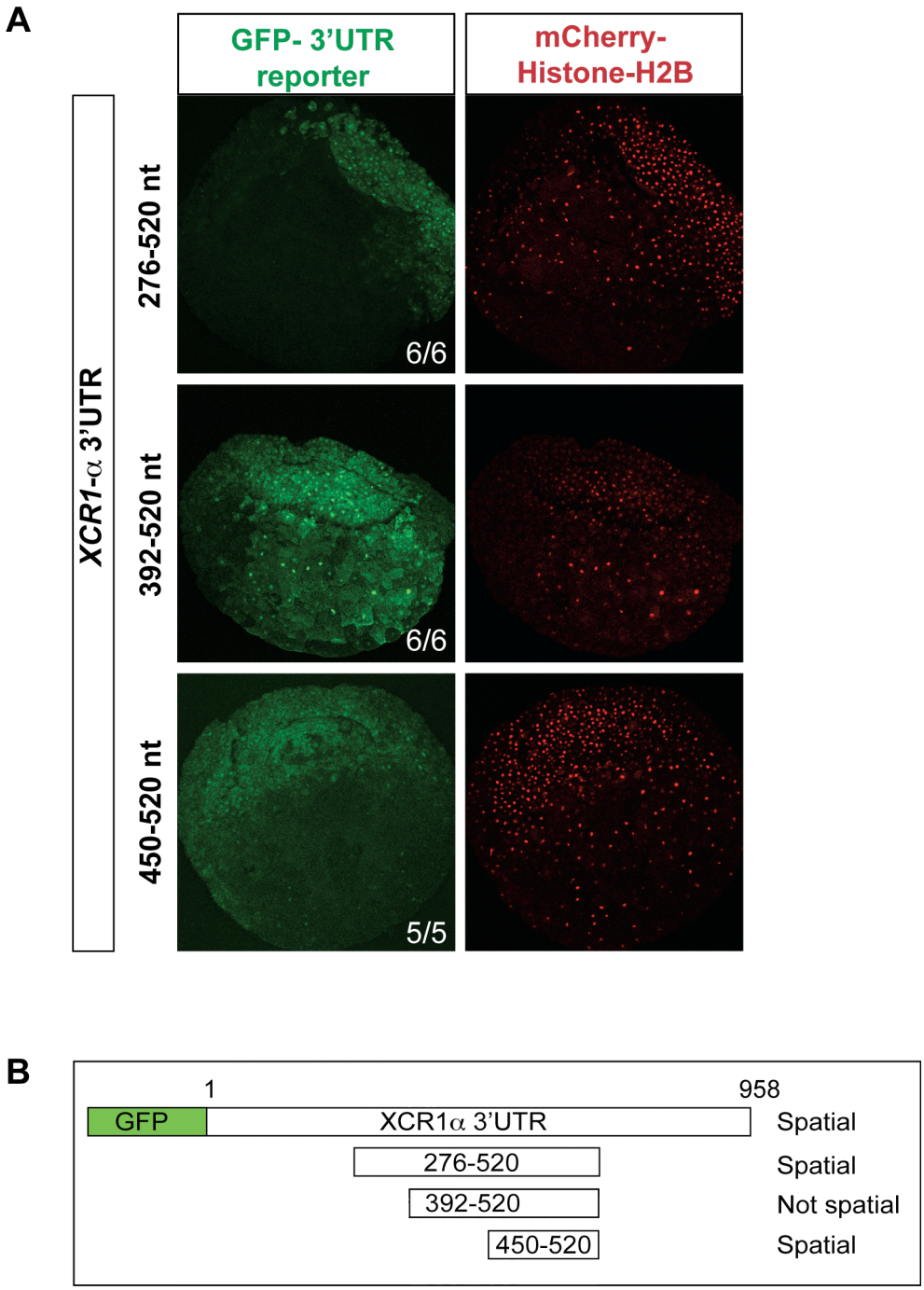
Figure 3.18 *GFP-XCR1α 276-402 3'UTR* reporter mRNA is degraded post-injection
20 μg of total RNA from embryos injected with full length *GFP-XCR1-α 3'UTR* reporter mRNA or a partial *GFP-276-402 3'UTR* reporter mRNA was analysed by RNAse protection assay using a *GFP* probe. *EF-1α* is used as a loading control. The protected fragments are indicated. Probes are shown prior to digestion. In the tRNA lane, no *Xenopus* embryo RNA was added.

The lack of GFP protein in the 276-402 clone was shown to be due to degradation of injected GFP-3'UTR mRNA, which is not suprising given that GFP with no 3'UTR is also unstable (Figure 3.18). This result indicates that there is an RNA-stabilising sequence between 402-461 to restore GFP expression. Next, the 276-520 minimal region of the *XCR1α* 3'UTR capable of conferring spatial regulation of GFP expression was subjected to finer mapping, shown in Figure 3.19A.

Since nucleotides 276-402 of the *XCR1α* 3'UTR were unstable, nucleotides 392-520 were cloned downstream of GFP. As shown in Figure 3.19, GFP expression was not spatially regulated when the 392-520 sequence was downstream of GFP. This fits with previous mapping data where nucleotides 392-958 conferred non-spatial expression of GFP (Figure 3.16) and the presence of an RNA-stabilising signal between nucleotides 402-461.

A GFP-partial 3'UTR reporter mRNA for nucleotides 450-520 of the *XCR1α* 3'UTR was also cloned, in order to exclude most of the RNA-stabilising signal present within nucleotides 402-461. The *GFP-XCR1α 450-520 3'UTR* reporter mRNA conferred weak but spatial expression of GFP. This suggests that this 70 nucleotide sequence is sufficient for spatial regulation, but another input is required to maintain higher GFP levels seen in the full length and 276-520 GFP-3'UTR reporter mRNAs.

In short, the minimal region sufficient for spatial regulation of the *XCR1α* 3'UTR at GFP levels comparable to the full length 3'UTR is the nucleotides 276-520. A sequence sufficient to confer spatial regulation of GFP expression at very low levels is located within nucleotides 450-520. An RNA-stabilising element is present between nucleotides 392-450. Since 392-958 nt of the *XCR1α* 3'UTR is not spatial, but 276-520 nt does confer spatial regulation, this suggests that there is also a spatial sequence element between 276-392 nt. This is illustrated in Figure 3.20 and a full summary of mapping results and analysis for the *XCR1α* 3'UTR is provided in Table 3.2.


A
GFP- 3'UTR reporter
mCherry-Histone-H2B
XCR1-α 3'UTR
276-520 nt
392-520 nt
450-520 nt
6/6
6/6
5/5
B
1
958
GFP
XCR1α 3'UTR
276-520
392-520
450-520
Spatial
Spatial
Not spatial
Spatial

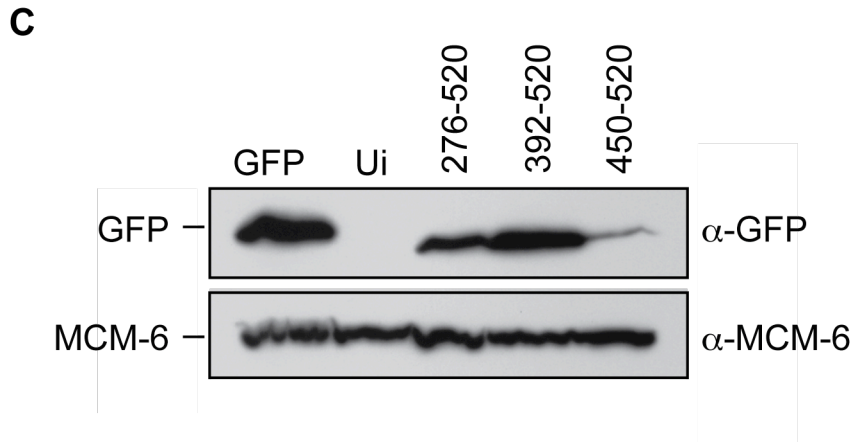


Figure 3.19 Fine mapping of the minimal region of the *XCR1α* 3'UTR shows that nucleotides 450-520 are sufficient for spatial regulation
A. Confocal microscopy assay for spatial regulation. *X. laevis* embryos were injected with GFP-partial 3'UTR reporter mRNAs containing nucleotides 276-20, 392-520 and 450-520 of the *XCR1-α* 3'UTR and analysed by cofocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is shown. mCherry-Histone-H2B expression is a control for spread of the injected mRNA. **B.** Schematic of partial GFP-*XCR1α* 3'UTR clones to map the spatial regulatory region within nucleotides 276-520 and summary of results. 'Spatial' indicates that the partial 3'UTR sequence conferred spatial regulation of GFP expression by confocal microscopy. **C.** Western blot showing GFP protein levels in stage 10 embryos injected with *GFP* mRNA, uninjected (Ui) or injected with GFP-*XCR1α* 3'UTR partial 3'UTR clones 276-520, 392-520 and 450-520. MCM-6 is used as a loading control. The result shown is an average of 20 embryos.

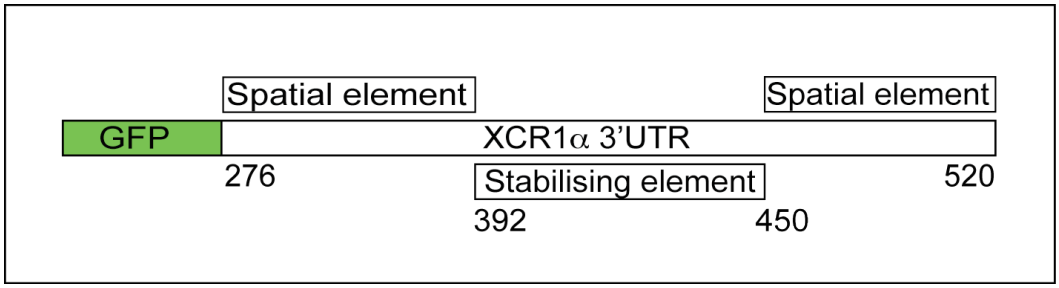


Figure 3.20 Mapping the spatial regulatory region of the *XCR1α* 3'UTR reveals three critical elements
Schematic of the minimal spatial regulatory region of the *XCR1α* 3'UTR, nucleotides 276-520. Fine mapping data indicates that there are two spatial elements and an RNA-stabilising element within this region.

GFP-3'UTR clone	Result	Analysis
XCR1α full length (1-958 nt)	Spatially regulated	The *XCR1α* 3'UTR confers spatial regulation of XCR1
XCR1α 1-150 nt	Weak and not spatially regulated	
XCR1α 1-200 nt	Not spatially regulated	
XCR1α 1-350 nt	Not spatially regulated	1-350 nt are not required for spatial regulation
XCR1α 286-637 nt	Spatially regulated	286-637 nt confer spatial regulation. Note that since 1-350 nt and 392-958 nt are not required for spatial regulation, this points to 351-392 nt containing a spatially regulating sequence
XCR1α 392-958 nt	Not spatially regulated	392-958 nt are not required for spatial regulation
XCR1α 276-402 nt	RNA is unstable	This sequence has no effect, as *GFP* mRNA with no 3'UTR is also unstable.
XCR1α 276-461 nt	Not spatially regulated	An RNA-stabilising site exists between nucleotides 402-461
XCR1α 276-520 nt	Spatially regulated	276-520 nt is the minimal region sufficient for spatial regulation. A sequence conferring spatial regulation is present between 462-520 nt in addition to another spatial element between 351-392 nt
XCR1α 276-579 nt	Not spatially regulated	An effect on mRNA secondary structure when adding 60 nt on to the spatially-regulated 276-520 nt reporter mRNA may explain this result
XCR1α 392-520 nt	Not spatially regulated	Includes an RNA-stabilising site between 402-461 which masks the effect of the 450-520 spatial element and fits with 392-958 nt as not required for spatial regulation
XCR1α 450-520 nt	Weak and spatially regulated	This sequence is sufficient to confer spatial regulation, but at very low protein levels
XCR1β 3'UTR 300-764 nt	Non-spatial GFP expression restored by deleting nt 1-299	Nucleotide 300 in the *XCR1β* 3'UTR is equivalent to nt 363 in the XCR1α 3'UTR when aligned. This result fits with *XCR1α* 392-958 nt being non-spatial.

Table 3.2 Summary of mapping the regulatory region of the *XCR1α* 3'UTR

3.2.5 Spatial regulation of XCR1 shows some Dicer dependency

Spatial regulation of XCR1 by the 3'UTR of *XCR1α* has been mapped to a critical region comprising nucleotides 276-520. Since the 3'UTRs of many genes are targeted by microRNAs, to determine whether spatial regulation mediated by the *XCR1α* 3'UTR could be a result of microRNA regulation, the microRNA-processing enzyme Dicer was targeted with an antisense morpholino to downregulate the production of mature microRNAs. If spatial regulation mediated by the *XCR1α* 3'UTR requires microRNAs, downregulation of Dicer should rescue spatial regulation of endogenous XCR1 or GFP expression in the vegetal pole when the *GFP-XCR1α 3'UTR* reporter mRNA is injected. Firstly, the efficacy of the translation-blocking Dicer morpholino was confirmed by fusing the morpholino binding site sequence upstream of the coding region of *GFP* (Gessert et al., 2010) and assessing the downregulation of GFP protein levels by Western blot (Figure 3.21A). Overexpression of 500 pg of morpholino binding site-GFP was inhibited by both 20 ng and 40 ng doses of the Dicer morpholino injected at the 1-cell stage. Injection of 20 ng of Dicer morpholino inhibited eye development and caused a loss of structures derived from neural crest cells such as cranial cartilage and melanocytes (Figure 3.21B). This confirms previous results obtained with this morpholino (Gessert et al., 2010).

Dicer knockdown had a limited rescue effect on spatial regulation mediated by the full length *XCR1α* 3'UTR as 1/6 morphant embryos showed partial rescue of GFP expression in the vegetal pole (Figure 3.22). Stronger rescue of spatial regulation was observed when Dicer morpholino was coinjected with a GFP-3'UTR reporter mRNA of the minimal 276-520 region of the 3'UTR. Embryos showed full rescue of GFP expression in the vegetal pole in 2/14 embryos, and partial rescue of GFP expression in 5/14 embryos (Figure 3.22). This suggests that spatial regulation mediated by the minimal spatial region of the *XCR1α* 3'UTR (nucleotides 276-520) shows some Dicer dependency and implicates microRNAs as candidate regulators for spatial regulation of XCR1.

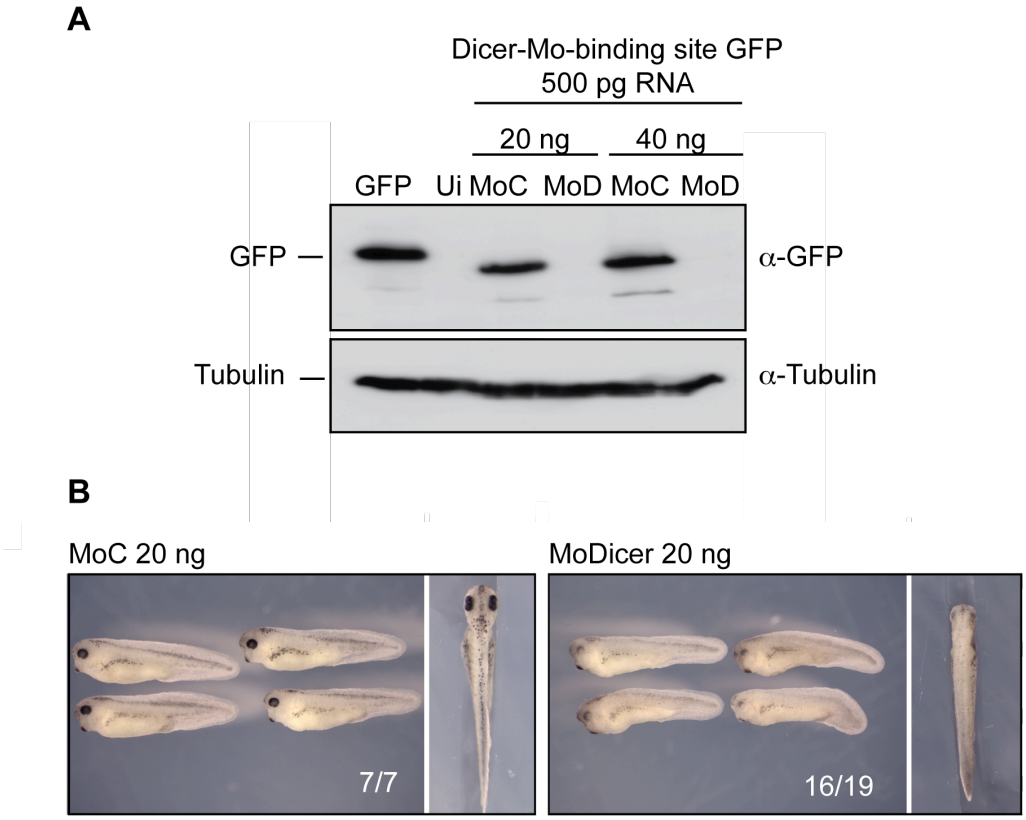


Figure 3.21 Morpholino knockdown of Dicer is effective and causes defective eye and neural crest development
A. Western blot for GFP in stage 9 *X. laevis* embryos injected with control GFP, uninjected (Ui) or injected with control morpholino (MoC) or Dicer morpholino (MoD) with 500 pg of Dicer-morphilino binding site GFP at the 1-cell stage. The Dicer morpholino is translation blocking and effectively suppresses Dicer-Mo-binding site GFP expression at doses of 20 ng and 40 ng. Tubulin is used as a loading control. **B.** *X. laevis* embryos injected with 20 ng of control morpholino or Dicer morpholino were assessed for phenotypic changes at stage 36. A lateral and dorsal view of morphant embryos is shown. The number of embryos with the phenotype shown out of the total number analysed is indicated. Knockdown of Dicer leads to inhibition of eye development and loss of neural crest derivatives such as craniofacial cartilage and melanocytes, as previously described (Gessert et al., 2010).

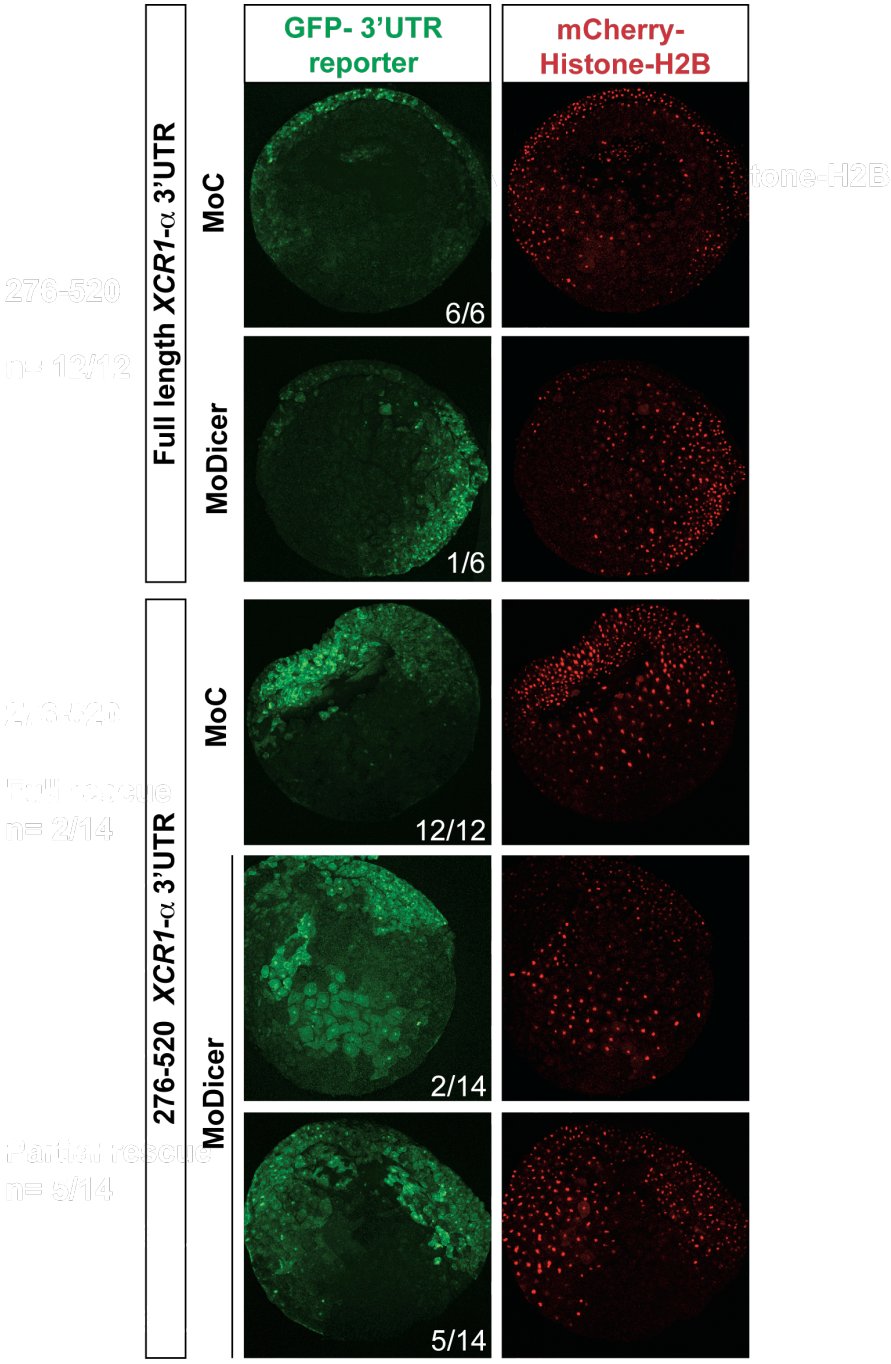


Figure 3.22 Dicer knockdown can rescue spatial regulation mediated by the 276-520 region of the *XCR1α* 3'UTR

Confocal assay for spatial regulation. *X. laevis* embryos were injected with 20 ng of control or Dicer morpholino with the full length *GFP-XCR1α 3'UTR* reporter mRNA or the GFP-276-520 3'UTR reporter mRNA and assayed for spatial regulation of GFP expression at stage 10. Dicer morphant embryos show weak partial rescue of vegetal GFP expression with the full length *XCR1α* 3'UTR (1/6 embryos). Dicer morphants display strong rescue of vegetal GFP expression compared to control morphants (2/14 embryos) and partial rescue in 5/14 morphants. The number of embryos with the GFP expression pattern shown out of the total analysed is indicated.

3.2.6 xla-miR-427 is a negative regulator of XCR1

As spatial regulation of XCR1 is conferred by the 3'UTR of *XCR1α* and partially Dicer-dependent, potential *Xenopus* microRNA binding sites in the *XCR1α* 3'UTR were predicted using the miRanda microRNA target prediction algorithm (Enright et al., 2003). This analysis was done in 2008, when miRbase was version 12. The known *Xenopus laevis* and *Xenopus tropicalis* microRNAs from miRbase (version 12) were scanned against the 3'UTR of *XCR1α* and a cut-off of a miRanda score of 120 and free energy of binding of -12 kcal/mol was taken. Since miRanda predicts microRNA binding sites using sequence-complementarity and free energy of miRNA-3'UTR binding to generate the miRanda score, with no bias towards complementarity of the microRNA seed sequence and the target 3'UTR, the miRanda results were also annotated for exact seed matches, defined as nucleotides 2-7 of the miRNA sequence matching perfectly to the *XCR1α* 3'UTR. 116 *Xenopus* microRNAs passed the miRanda score and free energy cut-off, of which the top ten hits sorted by miRanda score are presented in Table 3.3.

***Xenopus* microRNA**	**Total miRanda score**	**Total miRNA: 3'UTR binding energy (kcal/mol)**	**Site positions in *XCR1α* 3'UTR (nt)**	**Exact seed match position (nt)**
xla-miR-18	571	-89.29	850, 39, 746, 353	850, 39, 746
xla-miR-20	432	-84.5	45, 580, 265	45, 580, 265
xla-miR-427	400	-83.63	213, 40, 749	40
xla-miR-428	264	-79.26	213, 40	213, 40
xtr-let-7a	260	-69.46	144, 463	463
xtr-let-7b	248	-68.33	144, 455	None
xtr-let-7c	252	-63.55	144, 463	463
xtr-let-7e	247	-62.59	144, 460	None
xtr-let-7f	138	-62.29	144	None
xtr-let-7g	248	-62.21	94, 456	None

Table 3.3 Top ten miRanda scan hits for putative microRNA binding sites in the *XCR1α* 3'UTR sorted by miRanda score

To determine whether the microRNAs predicted to bind the *XCR1α* 3'UTR most favourably were responsible for spatial regulation of XCR1, the top three miRanda hits, xla-miR-18, xla-miR-20 and xla-miR-427 were downregulated in the *Xenopus laevis* embryo by morpholino injection. Efficacy of the xla-miR-427 morpholino, which is designed to prevent processing of the mature miRNA from the hairpin precursor (multi-blocker morpholino) was assessed by Northern blot for xla-miR-427 (Figure 3.23A). xla-miR-427 is expressed at stage 8, before mid-blastula transition and is therefore present maternally. Expression of mature xla-miR-427 persists until late neurula stages, whereas pre-miR-427 expression is curtailed post-gastrulation. Morpholino knockdown of xla-miR-427 is evident at stage 8, and most effective at stage 10. xla-miR-427 morphant embryos display a highly penetrant ventralised phenotype at stage 40 as previously described (Rosa et al., 2009), and secondary axis development is occasionally observed (Figure 3.23B).

Spatial regulation in microRNA morphants was assessed by Western blot for endogenous XCR1 protein in the animal cap, marginal zone and vegetal pole, and by GFP- *XCR1α* 3'UTR reporter assay. Morpholino knockdown of xla-miR-18, xla-miR-20 and xla-miR-427 had no effect on spatial regulation of XCR1 in either assay (data not shown for xla-miR-18 and 20), but knockdown of xla-miR-427 increased levels of XCR1 protein in the animal cap at stage 10 (Figure 3.24A). An increase in XCR1 levels was also evident in whole embryo protein extracts (Figure 3.24B). There was no effect of xla-miR-427 morpholino injection on spatial regulation mediated by the *XCR1α* 3'UTR when assayed by GFP-3'UTR reporter mRNA assay at stage 10 (Figure 3.25). These data suggest that xla-miR-427 is a negative, but not a spatial regulator of XCR1.

Next, the mechanism of xla-miR-427-mediated repression of XCR1 was investigated further. xla-miR-427-mediated repression is not mediated via the *XCR1α* 3'UTR, as GFP levels were comparable at stage 10 between control and xla-miR-427 morphant embryos which were coinjected with the *GFP- XCR1α 3'UTR* reporter mRNA (Figure 3.26A). xla-miR-427 does not affect *XCR1* transcript stability, as shown by RNAse protection assay in control and xla-miR-427 morphant embryos at stage 10 with a probe against the coding region of *XCR1* (Figure 3.26B). The Nodal ligands Xnr5 and Xnr6, which specifically signal through XCR1, are known xla-miR-427 targets (Rosa et al.,

2009). To test the possibility that XCR1 may be indirectly regulated by xla-miR-427 via Xnr5 and Xnr6 ligand levels, Xnr5 and Xnr6 were overexpressed at doses of 500 pg and XCR1 levels analysed by Western Blot (Figure 3.26C). If XCR1 levels were regulated by ligand availability, ectopic expression of Xnr5 and Xnr6 should have the same effect as the xla-miR-427 morpholino: increased XCR1 protein. As shown in Figure 3.26, XCR1 this was not the case, instead, overexpression of Xnr5 and Xnr6 appeared to decrease XCR1 expression.

In summary, xla-miR-427 is a negative regulator of XCR1 that does not target the *XCR1* 3'UTR or play a role in sensing XCR1-specific ligand levels in the Nodal pathway. However, it is still possible that xla-miR-427 directly targets the *XCR1* gene, as there is a perfect seed match site for xla-miR-427 in the coding region of XCR1.

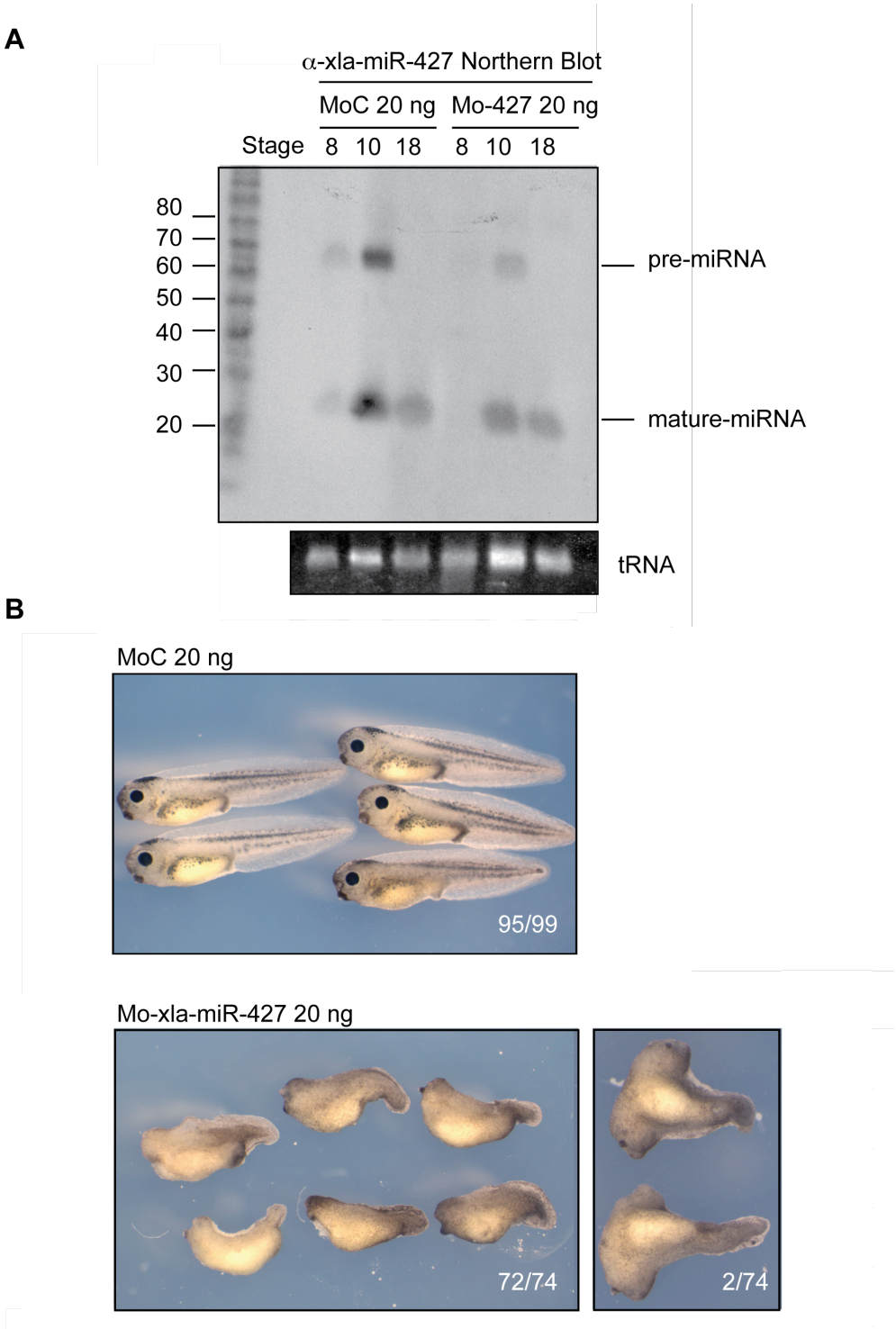


Figure 3.23 Morpholino knockdown of xla-miR-427 is effective at stages 8 and 10 and leads to a highly ventralised phenotype
A. Northern blot for xla-miR-427 in control (MoC) and xla-miR-427 morpholino (Mo-427) injected embryos. Ethidium bromide staining of tRNA is used as a loading control. **B.** xla-miR-427 morphants display a highly penetrant, ventralised phenotype compared to control morphants at stage 40. A secondary axis developed in 2/74 embryos. The number of embryos with the phenotype shown out of the total analysed is indicated.

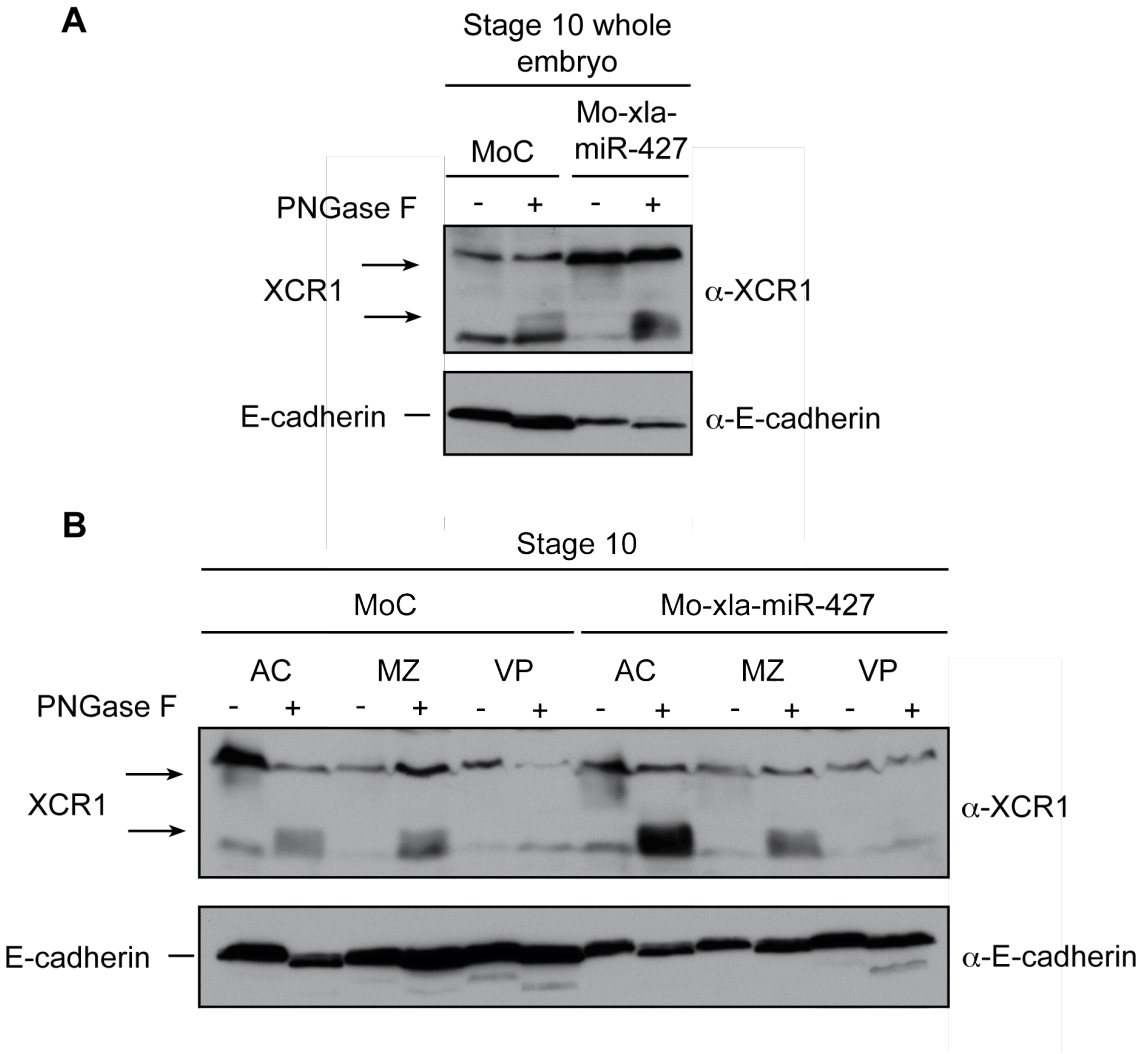


Figure 3.24 xla-miR-427 is a negative, but not a spatial regulator of XCR1
A. Western blot for XCR1. 20 ng of control (MoC) or xla-miR-427 morpholino (Mo-xla-miR-427) was injected at the 1-cell stage and *X. laevis* embryos were allowed to develop to stage 10. Whole embryo extracts were treated or not with PNGase F. E-cadherin is used a loading control because it is a membrane protein, as is XCR1. XCR1 levels are increased in xla-miR-427 morphant embryos compared to control morphants. **B.** Western blot for XCR1. Control and xla-miR-427 morphant embryos were dissected into animal caps (AC), marginal zones (MZ) and vegetal poles (VP) at stage 10 and protein extracts were blotted for XCR1. xla-miR-427 morphant embryos do not show any rescue of XCR1 expression in the vegetal pole, but XCR1 levels in the animal cap are increased.

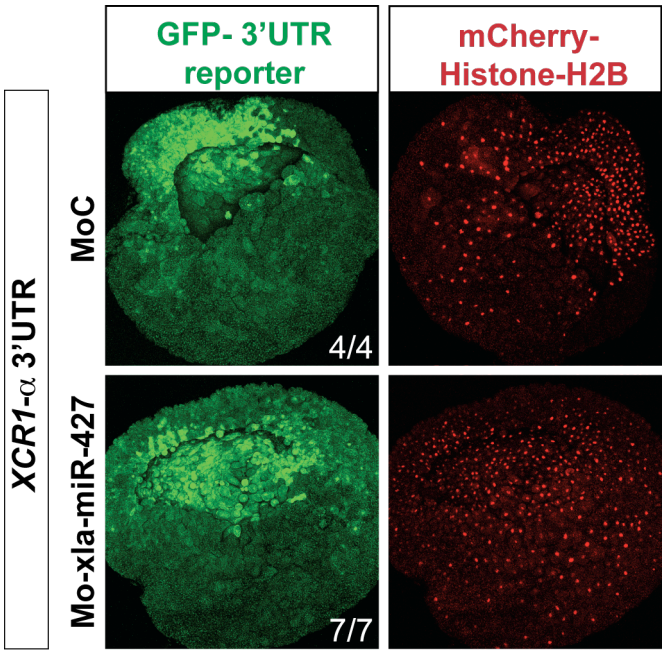


Figure 3.25 xla-miR-427 does not affect spatial regulation mediated by the *XCR1α* 3'UTR

Confocal microscopy assay for spatial regulation. *X. laevis* embryos were injected with *GFP-XCR1-α 3'UTR* reporter mRNA and either control morpholino (MoC) or xla-miR-427 morpholino and analysed by confocal microscopy at stage 10. The strong GFP expression in the xla-miR-427 morphant is at the back of the blastocoel. The total number of embryos with the GFP expression pattern shown out of the total analysed is indicated. *mCherry-Histone-H2B* mRNA was coinjected and is a control for spread of the injected mRNA.

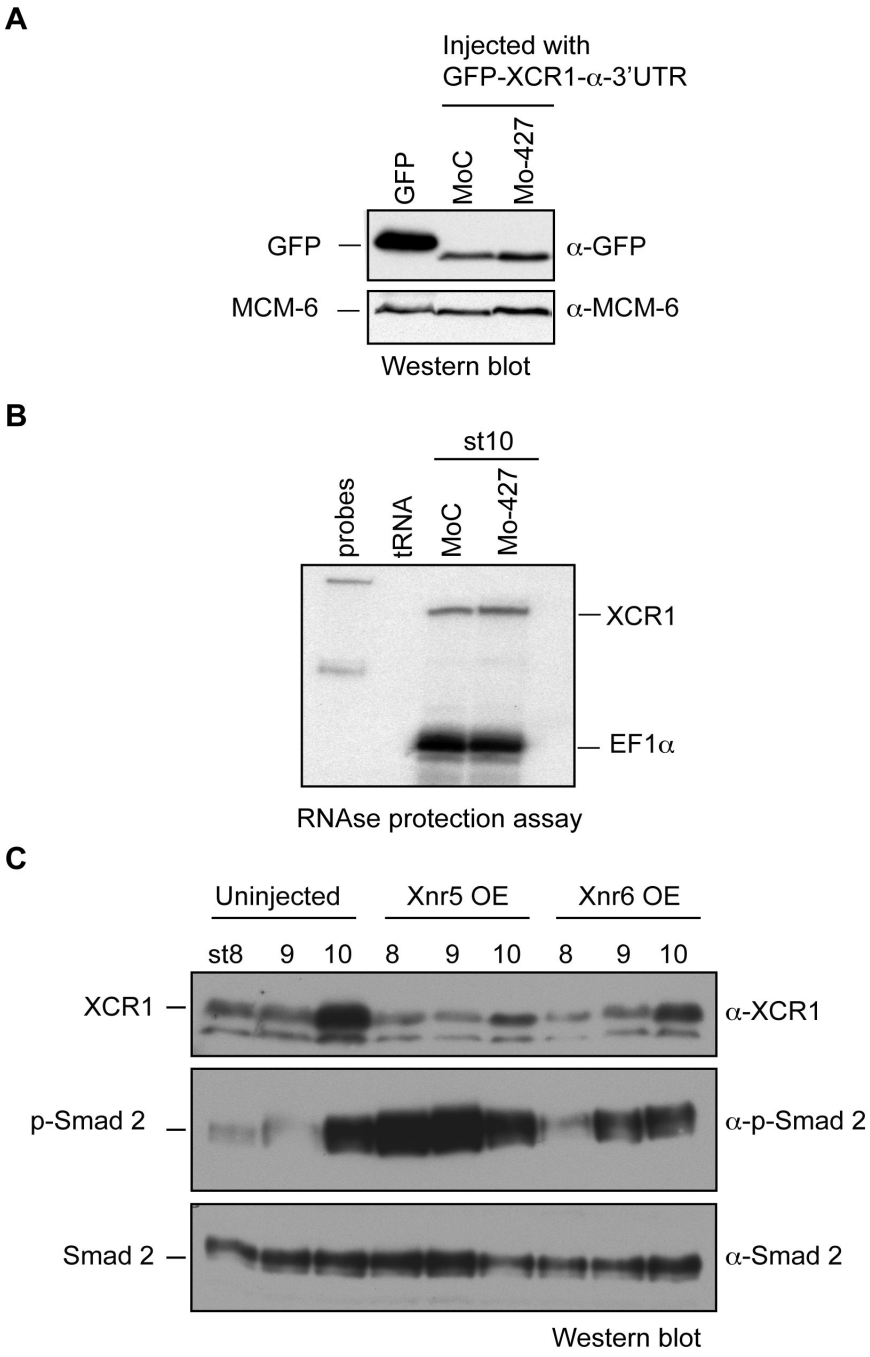


Figure 3.26 xla-miR-427 is an indirect negative regulator of XCR1
A. Western blot for GFP expression in stage 10 embryos injected with *GFP-XCR1α 3'UTR* reporter mRNA and 20 ng of control or xla-miR-427 morpholino. GFP overexpression is used as a control and MCM-6 serves as a loading control. **B.** 20 μg of total RNA from control or xla-miR-427 morphants at stage 10 was analysed by RNAse protection assay for *XCR1* mRNA levels with a probe against the *XCR1* coding region. *EF-1α* mRNA levels are used as a loading control. **C.** Western blots for phospho-Smad2 and XCR1 at stages 8, 9 and 10 in *X. laevis* embryos. Smad2 is used as a loading control. Embryos were uninjected or injected with 500 pg of *Xnr5* or *Xnr6* mRNA for overexpression (OE).

3.2.7 A mutagenesis approach to find spatial regulators

The critical region of the *XCR1α* 3'UTR for mediating spatial regulation of XCR1 was shown in section 3.2.4 to be nucleotides 276-520. Drawbacks of the miRanda microRNA target prediction approach described in section 3.2.6 include a lack of bias to exact seed matches in the target 3'UTR and simply taking all known *Xenopus* microRNAs in miRbase version 12 without the insight from expression data as to whether the microRNA is expressed in early development, and at the correct stages to potentially regulate XCR1: stage 8 and stage 10. Since limited numbers of *Xenopus* microRNAs have been identified and most have been found in adult tissues (Tang and Maxwell, 2008), very little was known about microRNA expression in early development. One study had identified 26 mature *Xenopus laevis* miRNA sequences expressed at stages from oocyte to tadpole (Watanabe et al., 2005).

In order to obtain an exhaustive list of microRNAs that were present in early development, and therefore could potentially regulate XCR1, I carried out Illumina genome-wide sequencing of small RNAs in *Xenopus tropicalis* at stages 8, 10 and 18. *Xenopus tropicalis* rather than *Xenopus laevis* small RNAs libraries were prepared because a complete reference genome has been sequenced for *Xenopus tropicalis* (Hellsten et al., 2010). The characterisation and analysis of the complete small RNA dataset is discussed extensively in Chapters 4 and 5. For the purposes of this chapter, the small RNA libraries of stages 8 and 10 were subjected to microRNA profiling. Small RNA sequences were compared to all microRNA sequences deposited in miRbase version 14, and microRNAs that were present at both stage 8 and stage 10 with a perfect seed match within the 276-520 nucleotide region of the *XCR1α* 3'UTR responsible for spatial regulation of XCR1 were retained as candidates. RNA-binding proteins with putative binding sites from a literature search within the 276-520 region were also noted. Figure 3.27 shows the predicted microRNA and RNA binding protein sites.

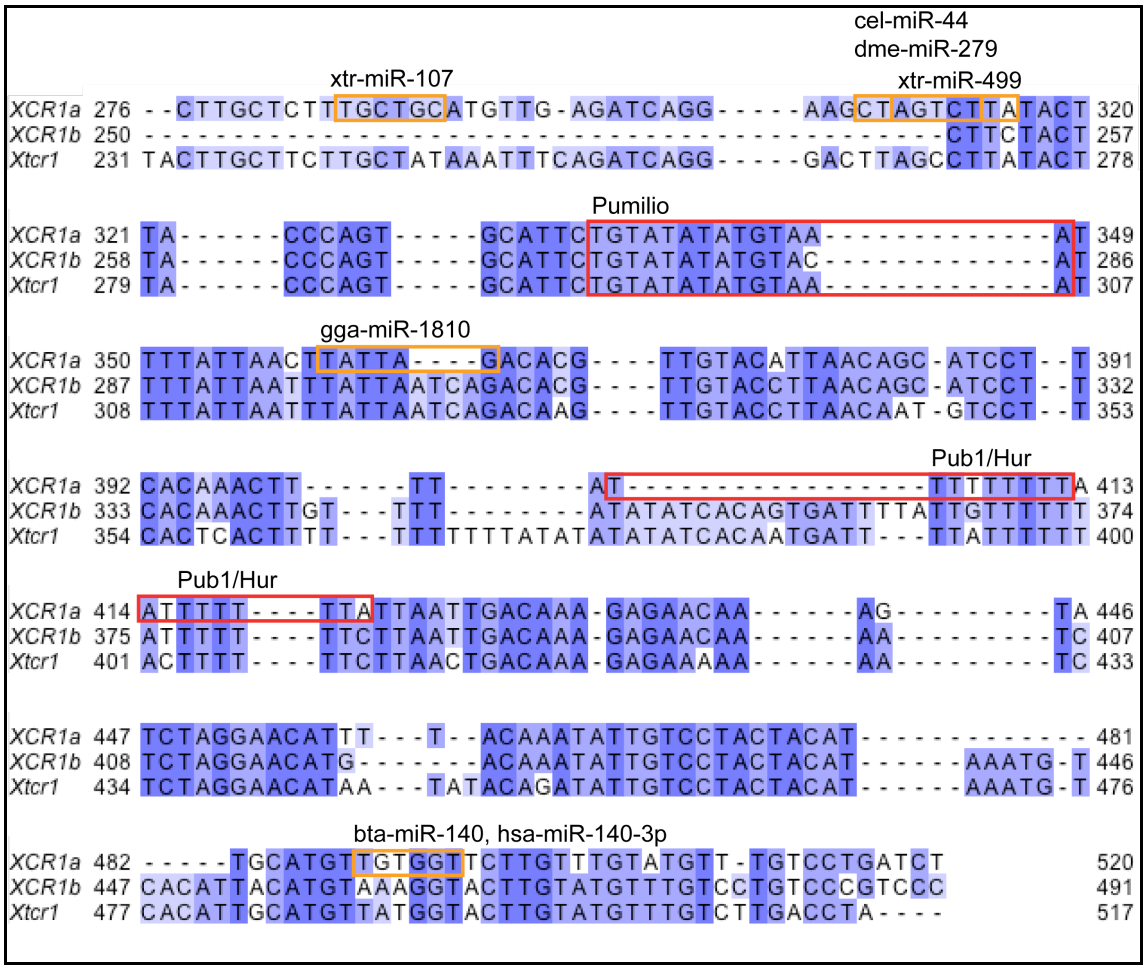


Figure 3.27 Candidate microRNA and RNA binding protein sites within the minimal region of the *XCR1α* 3'UTR

Xenopus XCR1 3'UTR sequences were aligned in Clustal W (Chenna et al., 2003). *XCR1α*, *XCR1β* and *Xtcr1* 3'UTR alignments to nucleotides 276-520 of the *XCR1α* 3'UTR are shown and coloured for percentage identity. Exact seed match sites for microRNAs that were sequenced in *Xenopus tropicalis* at both stage 8 and stage 10 in the 276-520 minimal spatial region of the *XCR1α* 3'UTR are shown in orange. Where an *X. tropicalis* microRNA was found in multiple species, only the xtr-miR label has been used for clarity. The name of the microRNA retains the name of the sequence used in miRbase version 14. Candidate RNA-binding protein sites are marked in red. Note that the microRNAs listed in this figure were obtained by allowing one mismatch between the small RNA sequencing read and the miRbase microRNA sequence.

The 276-520 region of the *XCR1α* 3'UTR contains predicted seed match sites for the following microRNAs sequenced in *Xenopus tropicalis*: bta-miR-107 and xtr-miR-107, overlapping sites for cel-miR-44, dme-miR-279, hsa-miR-499-5p and xtr-miR-499, gga-miR-1810 and a site for bta-miR-140 and hsa-miR-140-3p. The Pumilio family of RNA binding proteins recognise the motif TGTANA in the 3'UTR of target genes and are negative post-transcriptional regulators (Wharton and Aggarwal, 2006). There are two Pumilio sites in the 276-520 region. There are also overlapping sites for the yeast RNA binding protein poly(U) binding protein 1 (Pub1), which stabilises transcripts (Ruiz-Echevarria and Peltz, 2000) and the RNA-stabilising protein Hur (Lopez de Silanes et al., 2004). To determine whether these putative microRNA and RNA binding protein sites played a role in spatial regulation, GFP-mutant-276-520 nt- *XCR1α* 3'UTR reporters were cloned with multiple mutations in the predicted binding sites introduced by site-directed mutagenesis. The mutations are described in Table 3.4. Pumilio sites were analysed by antagonising the binding of Pumilio protein with a target protector oligo (Choi et al., 2007).

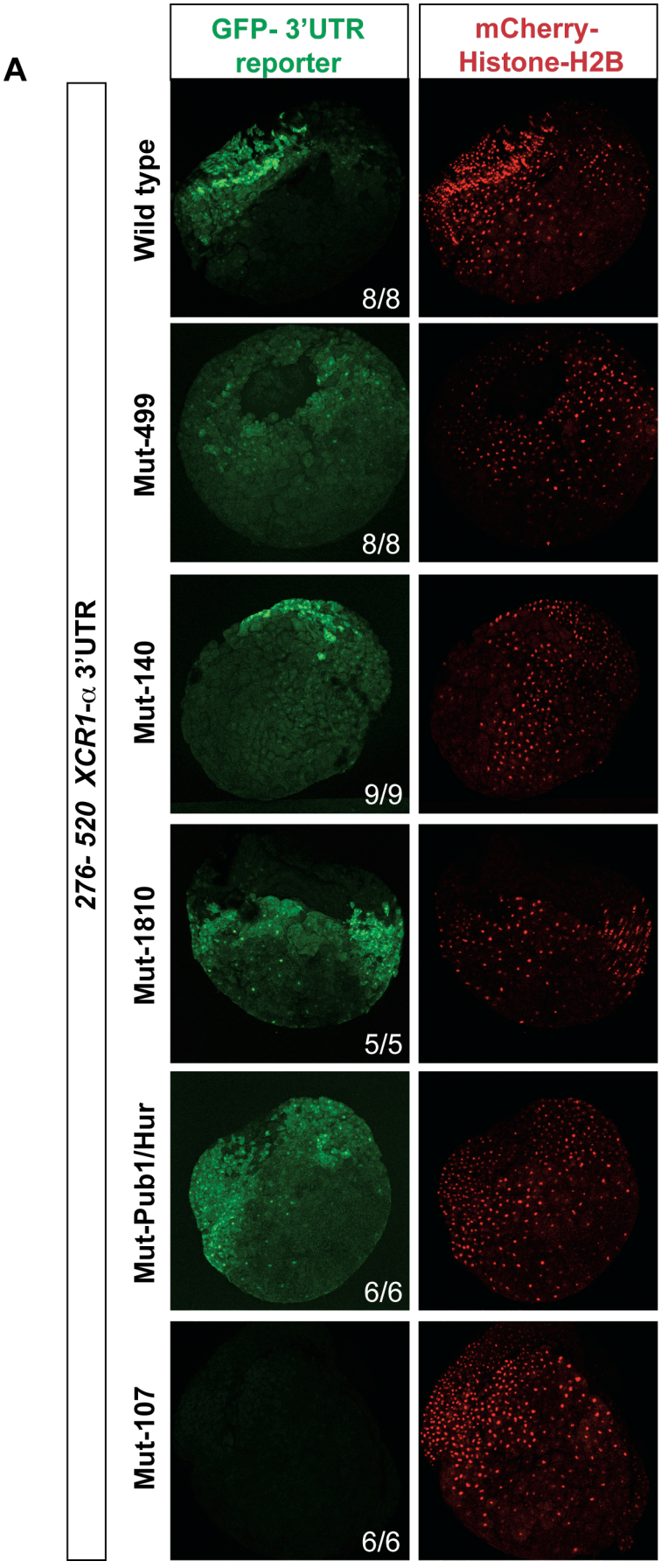
GFP-Mut-276-520-3'UTR reporter mRNAs for Mut-499, Mut-140, Mut-1810, Mut-Pub1/Hur and Mut-107 were analysed for their ability to confer spatial regulation of XCR1 by confocal microscopy and Western blot (Figure 3.28). In comparison to the wild-type 276-520 3'UTR, Mut-499, Mut-140 and Mut-Pub1/Hur had no effect on spatial regulation or GFP levels (Figure 3.28). Mut-1810, however, did revert spatial regulation as GFP was strongly expressed in the vegetal pole of stage 10 embryos, and total GFP levels were increased in the whole embryo. Mut-107 had no effect on spatial regulation, but strongly down-regulated GFP expression, as no GFP fluorescence was detectable by confocal microscopy, and only a trace of protein was detectable by Western blot.

Next, the effect of blocking the predicted Pumilio sites on spatial regulation was assessed (Figure 3.29). No effect of blocking Pumilio binding with 30 ng of target protecting oligo was seen on spatial regulation or GFP levels when coinjected with partial *GFP-XCR1α 3'UTR* reporter mRNAs (276-520, 286-637 and 276-402 to allow for concentration-dependent effects) or with the full length *GFP-XCR1β 3'UTR* where

the Pumilio sites are conserved. Spatial regulation is therefore independent of Pumilio proteins binding the *XCR1α* 3'UTR.

Candidate regulator	MicroRNA seed sequence	Predicted binding site in 3'UTR	Mutated binding site
bta-miR-107 xtr-miR-107	5'GCAGCA	5' TGCTGC	Mut-107 5'TCCACG
cel-miR-44 dme-miR-279	5' GACUAG	5'CTAGTC	Mut-499 5'CTTCAG
hsa-miR-499-5p xtr-miR-499	5'UAAGACU	5'AGTCTTA	Mut-499 5'TCAGTAA
gga-miR-1810	5'CUAAUA	5'TATTAG	Mut-1810 5'ATATAC
bta-miR-140 hsa-miR-140-3p	5' ACCACA	5'TGTGGT	Mut-140 5'AGTCCA
Pub1		5'[TTTTTTTTT]-A[ATTTTTTT]	**Mut-Pub1/Hur** 5'[TATGAACTT]A[ATCAATAC]
Hur		5'[TTTTATTTT-TTTTTAATTTT]	**Mut-Pub1/Hur** 5'[TTTTATATGAACTTAATCAAT]AC
Pumilio		5'[TGTATA]TA[-TGTAAA]	Target protector over binding site

Table 3.4 MicroRNA and RNA-binding protein binding site manipulations in the *XCR1α* 3'UTR 276-520 3'UTR sequence
Nucleotides mutated by site directed mutagenesis are coloured red. Square brackets define the limits of RNA binding protein binding sites. Mut-499 disrupts the predicted binding sites for cel-miR-44, dme-miR-279, hsa-miR-499-5p and xtr-miR-499 shown by yellow shading. Mut-Pub1/Hur1 disrupts the predicted binding for both Pub1 and Hur.


A
GFP- 3'UTR reporter
mCherry-Histone-H2B
276- 520 XCR1-α 3'UTR
Wild type
8/8
Mut-499
8/8
Mut-140
9/9
Mut-1810
5/5
Mut-Pub1/Hur
6/6
Mut-107
6/6

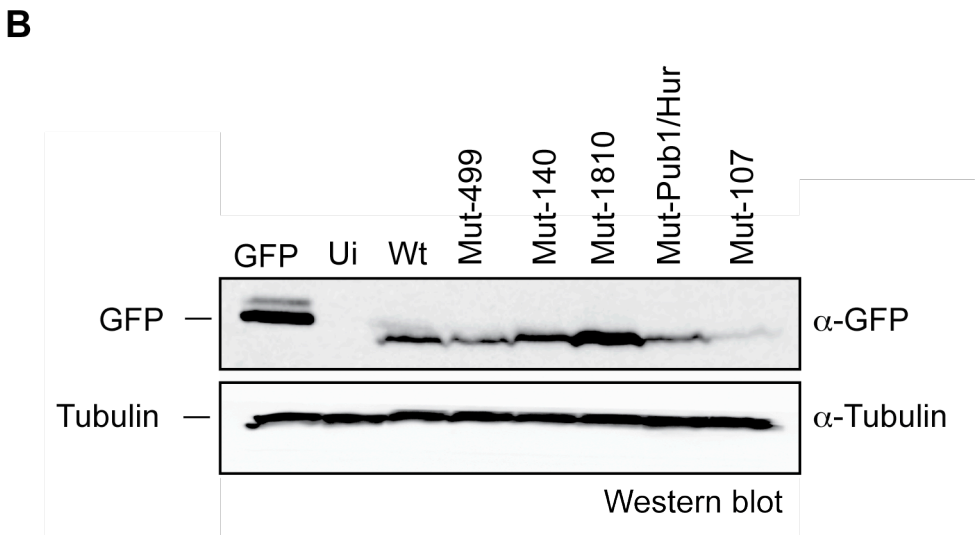


Figure 3.28 *GFP-Mut-276-520 3'UTR* reporter mRNA analysis shows that Mut-1810 reverts spatial regulation at stage 10
A. Confocal microscopy assay for spatial regulation. *X. laevis* embryos were injected with wildtype or *GFP-Mut-276-520 3'UTR* reporter mRNAs and *mCherry-Histone-H2B* mRNA and analysed by confocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is indicated.
B. Western blot for GFP. Whole embryo extracts were prepared from stage 10 embryos injected with control GFP, uninjected (Ui) or injected with wildtype (Wt) or mutant (Mut) *GFP-276-520 3'UTR* reporter mRNAs as described in Table 3.4. Tubulin serves as a loading control.

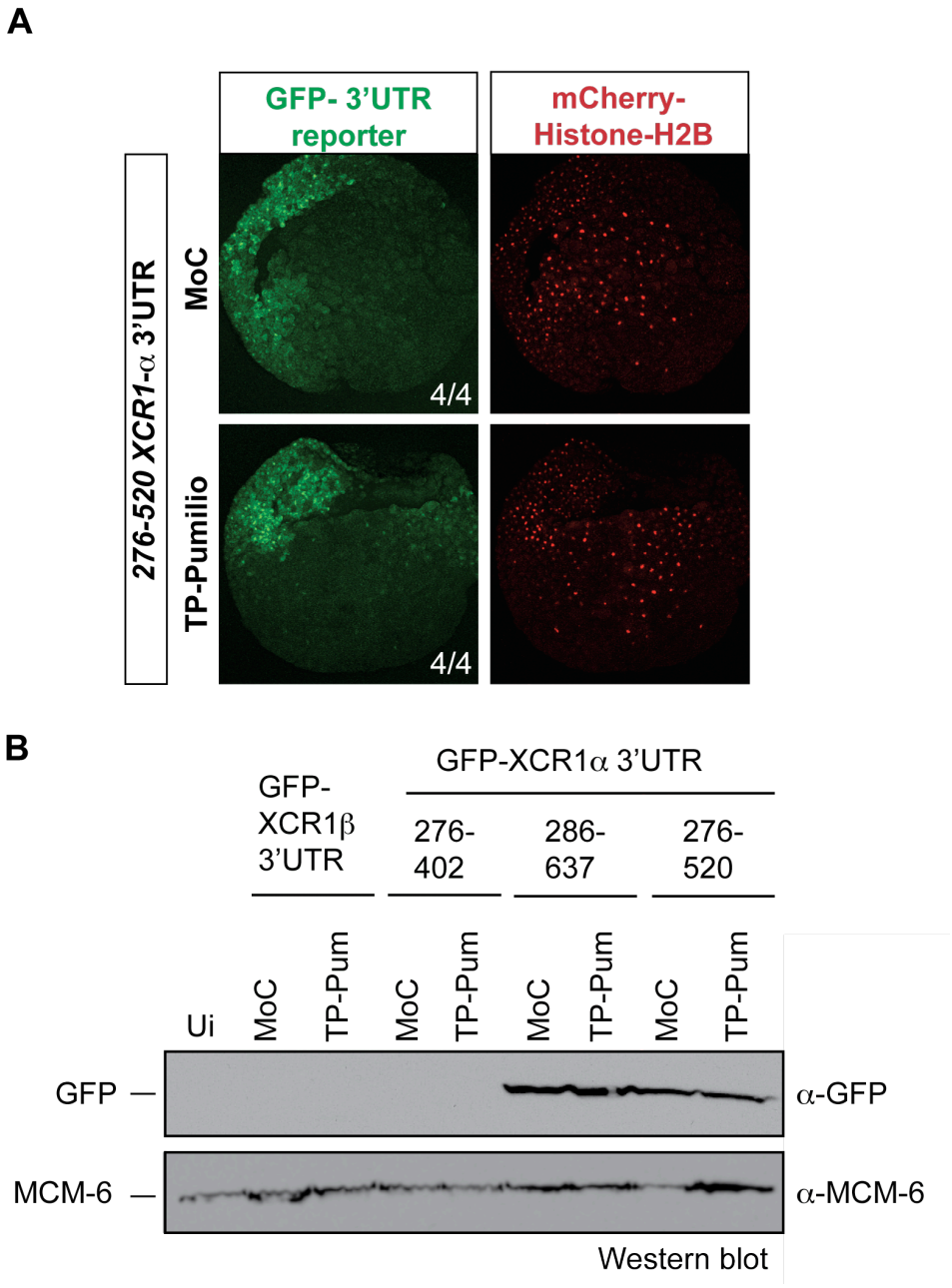


Figure 3.29 Spatial regulation is not dependent on Pumilio proteins
A. Confocal microscopy assay for spatial regulation. *X. laevis* embryos were injected with the miminal spatial region *GFP-3'276-520 3'UTR* reporter mRNA and 30 ng of control morpholino (MoC) or 30 ng of a target protector against Pumilio binding sites (TP-Pumilio) and *mCherry-Histone-H2B* mRNA and analysed by confocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is indicated. **B.** Western blot for GFP. Whole embryo extracts were prepared from stage 10 embryos injected with control GFP, uninjected (Ui) or injected with GFP-3'UTR reporter mRNAs and control morpholino or target protector against Pumilio sites (TP-Pum). MCM-6 serves as a loading control.

In order to determine if the Mut-1810 and Mut-107 binding site mutations had the same effect on spatial regulation or GFP levels in the context of the full length *XCR1α* 3'UTR sequence as the 276-520 nt region, these mutations were introduced into the full length 3'UTR. *GFP-Mut-1810- XCR1α* 3'UTR and *GFP-Mut-107 XCR1α* 3'UTR reporter mRNAs were analysed for spatial regulation (Figure 3.30). Mut-1810 reverted spatial regulation in the context of the full length 3'UTR, shown by rescue of GFP expression in the vegetal pole, accompanied by an large increase in GFP levels shown by Western blot. Mut-107 did not have the effect of decreasing GFP expression to undetectable levels in the context of the full length sequence, but instead increased GFP expression, although no striking effect on spatial regulation (as in Mut-1810) was evident. In short, spatial regulation mediated by the *XCR1α* 3'UTR can be abolished by mutating 5' TATTAG to 5' ATATAC beginning at nucleotide 360 in the 3'UTR. This sequence is an exact seed match for gga-miR-1810.

The microRNA gga-miR-1810 was identified in the stage 8 and stage 10 *Xenopus tropicalis* small RNA libraries at very low read numbers: 1 read at stage 8, 3 reads in the stage 10 whole embryo library, 14 reads in the stage 10 animal pole library and 15 reads in the stage 10 vegetal pole library. To validate the expression of gga-miR-1810 in both frog species, I performed microRNA-qPCR and Northern blot analysis. gga-miR-1810 was detectable by qPCR in both *Xenopus tropicalis* and *Xenopus laevis* at stages 8, 10 and 18 (Figure 3.31). The qPCR profile of gga-miR-1810 in *X. tropicalis* was similar to the small RNA-seq read number profile, whilst gga-miR-1810 expression decreases specifically at stage 10 in *X. laevis*. Mature gga-miR-1810 was not detectable by Northern blot analysis (expected size 20 nt) (Figure 3.32), however, a pre-miRNA band of 70 nt was observed. The lack of a mature miRNA band could be due to the low abundance of gga-miR-1810 indicated by the low small RNA-seq read numbers and the lower sensitivity of Northern blotting in comparison to qPCR (see Chapter 4). Next, gga-miR-1810 was inhibited by antagomir antisense oligo (Ambion) injection in *X. laevis* and spatial regulation of the GFP-*XCR1α* 3'UTR reporter was analysed. Antagomirs are cholesterol-conjugated small RNAs with a modified phosphodiester backbone and a 2'-*O*-methyl modifcation on ribose (Krützfeldt et al., 2006). Antagomir doses of 0.2 pmol, 0.5 pmol, 1 pmol, 2 pmol and 4 pmol were tested. Embryos injected

with 1-4 pmol of negative control antagomir did not develop normally in comparison to uninjected embryos, suggesting that this dose is toxic. No effect on spatial regulation was observed at lower doses of gga-miR-1810 antagomir (Figure 3.33). The mature sequence of gga-miR-1810 aligns to *Xenopus tropicalis* 28S ribosomal RNA, and gga-miR-1810 has been re-annotated as a ribosomal RNA fragment in the latest version of miRbase (version 17). Moreover, when the mature gga-miR-1810 sequence is aligned to the *Xenopus tropicalis* genome (14 alignment sites) and the surrounding sequence extracted and folded, no hairpin structures pass basic criteria for a microRNA pre-miRNA structure, such as the mature sequence located in one arm of the hairpin. This suggests that although this small RNA sequence does exist in both frog species, it is probably not a microRNA and clearly does not affect spatial regulation of XCR1 mediated by the *XCR1α* 3'UTR, as shown by antagomir inhibition. Nevertheless, the region of the *XCR1α* 3'UTR necessary for spatial regulation has been mapped to four nucleotides and potential regulators binding this site will be discussed in section 3.3.

3.2.7.1 Integration of 3'UTR mapping and mutagenesis data

The data presented have shown that the 5' TATTAG sequence beginning at nucleotide 360 in the *XCR1α* 3'UTR is necessary for spatial regulation, as when it is mutated to 5' ATATAC, spatial regulation is reverted in the context of nucleotides 276-520 and the full length 3'UTR. The minimal region of the 3'UTR sufficient to recapitulate spatial regulation at GFP levels similar to the full length 3'UTR is nucleotides 276-520. When this minimal region is further dissected, a *GFP-3'UTR* mRNA containing nucleotides 276-402 is unstable, as is GFP with no 3'UTR, and is stabilised by the addition of nucleotides 403-461. The 276-461nt GFP-reporter mRNA does not confer spatial regulation of GFP expression. This is an apparent contradiction, as this sequence includes the 5'TATTAG sequence at 360 nt, and therefore should confer spatial regulation if this sequence was both necessary and sufficient. However, a sequence downstream of 461 nt (i.e. within 462-520 nt) is also required. Further mapping revealed that nucleotides 450-520 alone conferred very weak spatial regulation of GFP expression and thus contain the second sequence necessary for spatial regulation. This sequence combined with the RNA-stabilising sequence in the 392-520 nt GFP-3'UTR reporter does not confer spatial regulation, therefore all three sequences

are required: the 5' TATTAG at 360 nt, the RNA stabilising sequence at 403-450 nt and the 450-520 nt sequence. This model will be presented in the section 3.3.

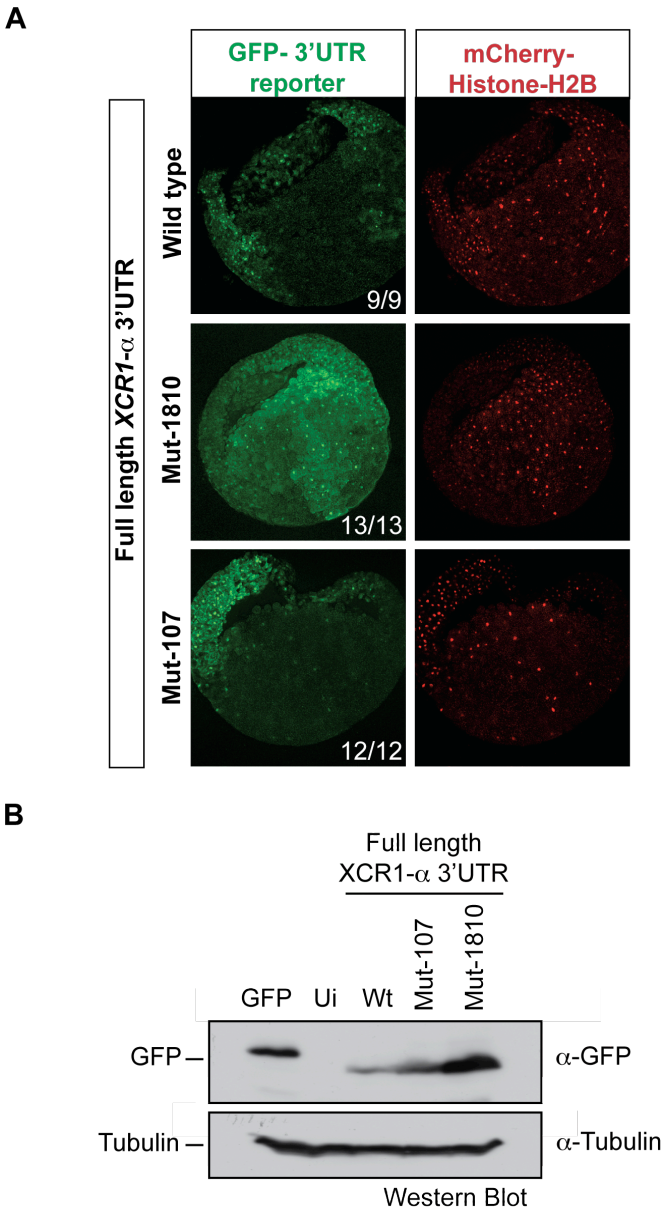


Figure 3.30 Mut-1810 reverts spatial regulation in the context of the full length *XCR1α* 3'UTR

A. Confocal microscopy assay for spatial regulation. *X. laevis* embryos were injected with wildtype or mutated (Mut-1810, Mut-107) full length *GFP-XCR1α 3'UTR* reporter mRNA and *mCherry-Histone-H2B* mRNA and analysed by confocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is indicated. **B.** Western blot for GFP. Whole embryo extracts were prepared from stage 10 embryos injected with control GFP, uninjected (Ui) or injected with wildtype or mutated (Mut-1810, Mut-107) full length *GFP-XCR1α 3'UTR* reporter mRNA. Tubulin serves as a loading control.

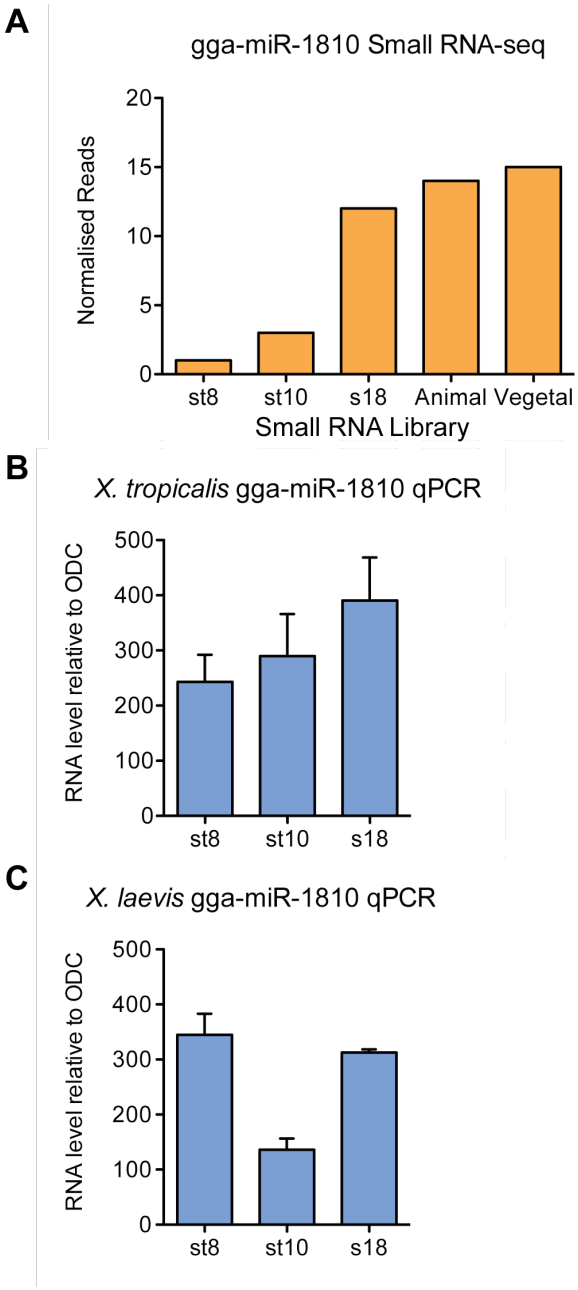


Figure 3.31 Ratification of gga-miR-1810 in *Xenopus tropicalis* and *Xenopus laevis*
A. Small RNA-seq reads profile for gga-miR-1810 in *X. tropicalis* small RNA libraries at stages 8, 10, 18 and stage 10 animal and vegetal libraries. **B.** gga-miR-1810 qPCR profile in *X. tropicalis.* Total RNA was isolated from stage 8, stage 10 and stage 18 embryos and prepared for microRNA-qPCR (see Figure 4.19). gga-miR-1810 levels were normalised to *ODC* mRNA levels and error bars are standard deviations of triplicate repeats **C.** gga-miR-1810 qPCR profile in *X. laevis* at stages 8, 10 and 18.

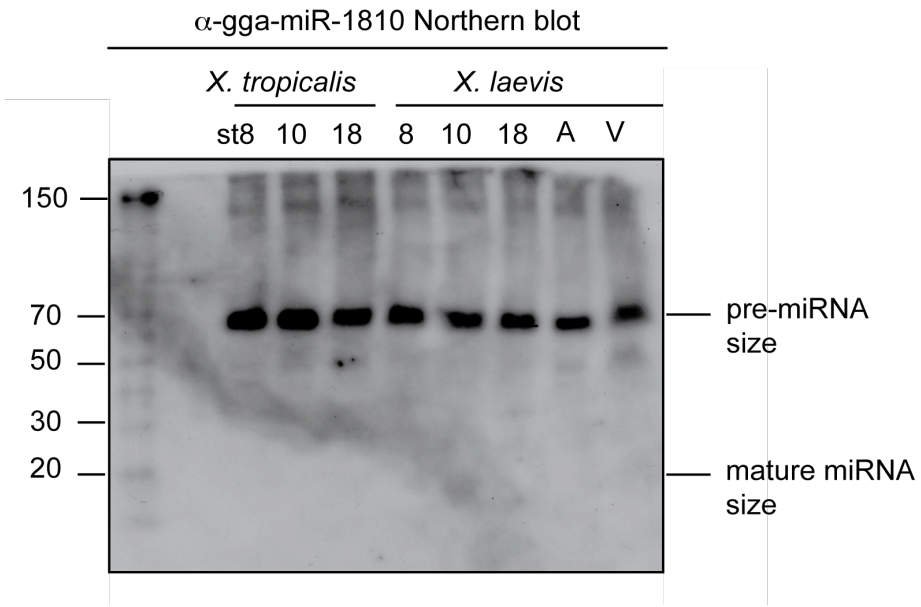


Figure 3.32 gga-miR-1810 Northern blot shows a pre-miRNA precursor
X. tropicalis and *X. laevis* total RNA from whole embryos or stage 10 animal (A) and vegetal (V) halves was probed for gga-miR-1810. Mature gga-miR-1810 is 20 nucleotides long from the small RNA-seq data and the pre-miRNA hairpin is predicted to be 77nt according to miRbase. A band of the correct size is evident for the gga-miR-1810 pre-miRNA, but not the mature miRNA.

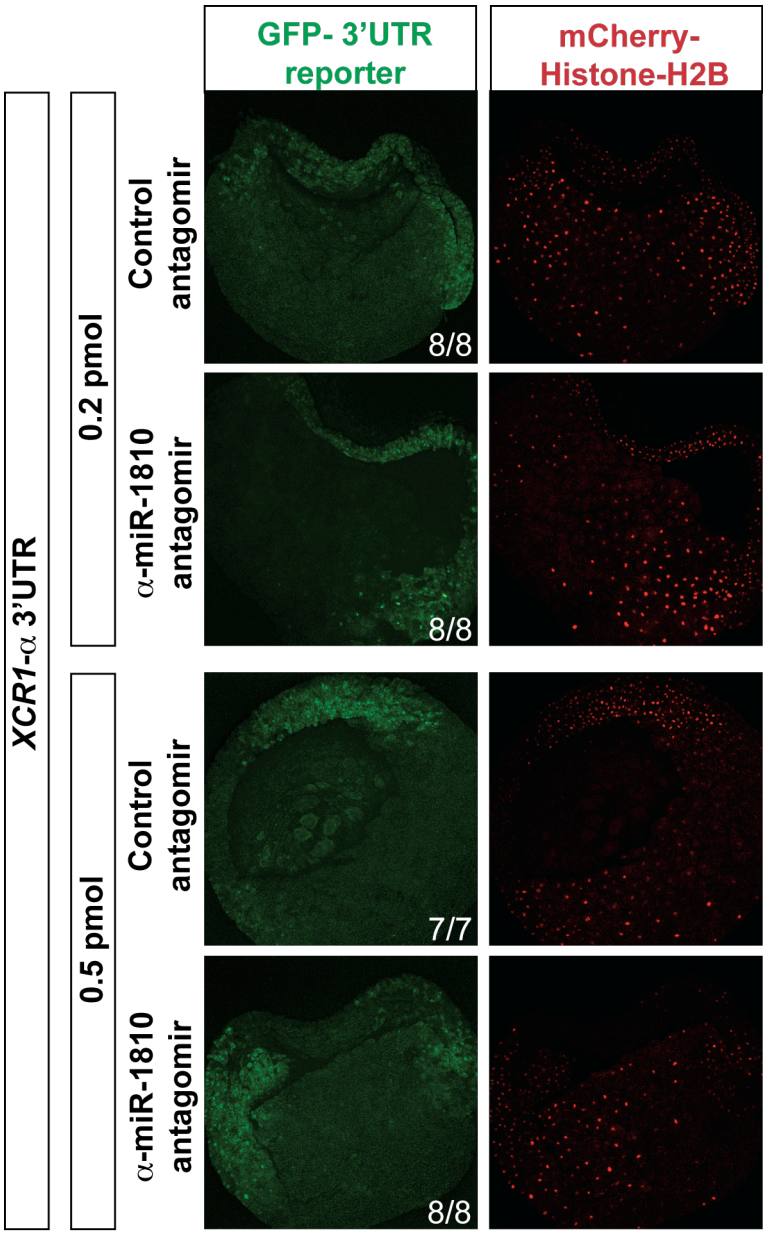


Figure 3.33 Knockdown of gga-miR-1810 does not affect spatial regulation of GFP-*XCR1α* 3'UTR

Confocal assay for spatial regulation. *X. laevis* embryos were injected with full length *GFP-XCR1α 3'UTR* reporter mRNA, control or α-gga-miR-1810 antagomir at 0.2 pmol or 0.5 pmol doses and *mCherry-Histone-H2B* mRNA. Embryos were analysed by confocal microscopy at stage 10. The total number of embryos with the GFP expression pattern shown out of the total analysed is indicated.

3.3 Discussion

3.3.1 Summary of results

- Spatial regulation of XCR1 is required to maintain the correct timing and magnitude of Nodal signalling during mesoderm induction via phospho-Smad2
- xla-miR-427 is an indirect negative regulator of XCR1
- Spatial regulation of XCR1 shows some Dicer-dependency and is mediated through the 3'UTR of *XCR1α*
- Nucleotides 276-520 define the minimal region of the *XCR1α* 3'UTR sufficient for spatial regulation
- There are three sequence elements required for spatial regulation: a necessary spatial site at nt 360, an RNA stabilising site at 402-461 and a spatial site at 450-520 nt

3.3.2 Spatial regulation of Nodal signalling by XCR1

The biological role of spatial regulation of XCR1 can be illuminated by replacing XCR1 protein in the vegetal pole. Embryos with vegetal XCR1 showed premature onset of Smad2 phosphorylation at stage 9, reduced mesodermal markers and upregulation of the Nodal inhibitor Lefty at stage 10 and dorsalisation and defects in head development at stage 32. Nodal signalling in the vegetal pole originates from induction of Xnr expression by the transcription factor VegT, which is maternally deposited in the vegetal pole (Zhang and King, 1996). Individual Nodal ligands require different XCR coreceptors to signal (Dorey and Hill, 2006). Xnr5 specifically requires XCR1, whereas Xnr1, Xnr2, Xnr4 and Xnr5 require both XCR1 and XCR3 to signal. When XCR1 is overexpressed vegetally, signalling by Xnr5 is introduced to the vegetal pole as XCR1 protein is normally absent from the vegetal pole. Signalling through Xnr1, Xnr2, Xnr4 and Xnr6 is potentiated, as these ligands require both XCR1 and XCR3 to signal (Dorey and Hill, 2006). This results in increased levels of phospho-Smad2 and consequent induction of *Lefty*, a Nodal target gene (Saijoh et al., 2000) (Figure 3.34). Since the effect of XCR1 replacement is on mesodermal markers, rather than in the prospective endoderm where XCR1 is added, a cell-non-autonomous factor must be responsible.

Lefty is a diffusible inhibitor of Nodal signalling that competes with Nodal for binding to XCR1 (and other EGF-CFCs) to form a negative feedback loop (Meno et al., 1999). *Lefty* mRNA is specifically upregulated in the prospective mesoderm/marginal zone as detected by *in situ* hybridisation in embryos with XCR1 vegetally overexpressed and could thus explain the downregulation of mesodermal markers *Xbra*, *Forkhead* and *Goosecoid* at stage 10. Later phenotypes of XCR1 vegetally replaced embryos are consistent with a lack of trunk and head mesoderm. One study has shown that Lefty overexpression inhibits mesoderm formation (shown by *Xbra* downregulation) but also causes gastrulation defects (Tanegashima et al., 2000). It could be that the endoderm is less sensitive to levels of Lefty than mesoderm, thus explaining the lack of an effect on endoderm markers. In short, spatial regulation of XCR1 is critical to maintain the timing and spatial control of Nodal signalling during mesoderm specification in gastrulation.

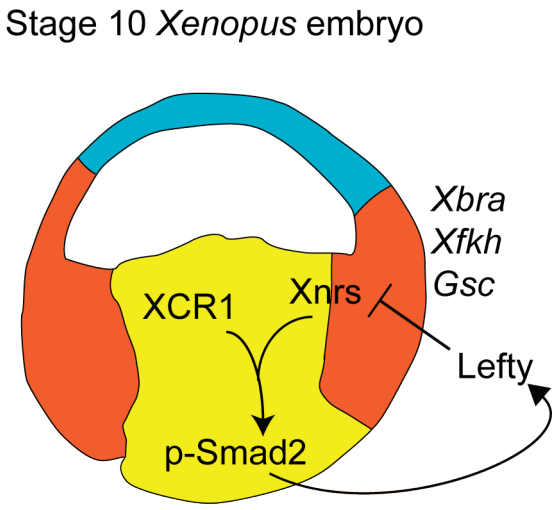


Figure 3.34 XCR1 replacement in the vegetal pole restricts mesoderm formation via increased levels of the Nodal pathway inhibitor, Lefty
Targeted overexpression of XCR1 to the vegetal pole potentiates Nodal signalling through Xnr1, Xnr2, Xnr4 and Xnr5 to increase phosphorylation of Smad2. The Nodal target gene Lefty is upregulated by high phospho-Smad2 levels and acts cell non-autonomously to inhibit Nodal signalling in the prospective mesoderm, thus downregulating mesodermal markers *Xbra*, *Xforkhead* and *Goosecoid*.

I have shown that XCR1 is negatively regulated by the microRNA xla-miR-427. The mechanism of xla-miR-427 regulation of XCR1 is unclear, but is independent of the

3'UTR and levels of Xnr5 and Xnr6. There is however, a perfect xla-miR-427 seed match site in the *XCR1* coding region that could be investigated as a potential regulatory site by a mutagenesis approach. It has been previously shown that xla-miR-427 also modulates Nodal signalling by 3'UTR-mediated repression of *Lefty*, *Xnr5* and *Xnr6b* (Rosa et al., 2009). Loss of function analysis shows that xla-miR-427 is required for proper mesoderm induction and organiser formation (Rosa et al., 2009). xla-miR-427 therefore modulates Nodal signalling at multiple levels in the pathway: by inhibition of ligand and inhibitor expression and indirect downregulation of XCR1 coreceptor expression. The inputs to phospho-Smad2 levels in Nodal signalling through XCR1 are summarised in Figure 3.35. The zebrafish homologue of xla-miR-427, miR-430 was also shown to inhibit *Lefty* and the Nodal ligand *Squint* (Choi et al., 2007). The human homologue, miR-302, also inhibits *Lefty* in human embryonic stem cells (Barroso-Deljesus et al., 2011; Rosa et al., 2009). Further conservation of xla-miR-427/miR-430 function has also been demonstrated as miR-427/miR-430 initiate degradation of maternal mRNAs after zygotic transcription begins by deadenylation of target mRNAs in both *Xenopus* (Lund et al., 2009) and zebrafish (Giraldez et al., 2006). The Nodal pathway is also subjected to microRNA regulation in *Xenopus* by xla-miR-15a and xla-miR-16 targeting of the *ActRIIA* Nodal Type II receptor (Martello et al., 2007).

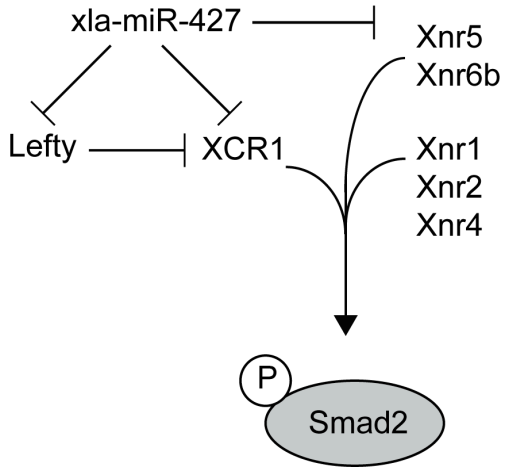


Figure 3.35 xla-miR-427 regulation of Nodal signalling via controlling inputs to phospho-Smad2 levels
xla-miR-427 represses both Xnr ligands and the Nodal inhibitor Lefty, which competes with Xnrs for XCR1 binding to antagonise Nodal signalling. xla-miR-427 also decreases XCR1 protein levels via a 3' UTR-independent mechanism. Phospho-Smad2 levels integrate these signal inputs to activate Nodal target genes.

3.3.3 Spatial regulation of XCR1 requires a minimal region of the 3'UTR containing 3 key elements and is partially Dicer-dependent

As a possible additional level of microRNA modulation of Nodal signalling, spatial regulation of XCR1 shows some Dicer dependency and is dependent on the 3'UTR, suggesting that microRNAs may directly regulate spatial control of Nodal signalling through XCR1. A model of the regulation within the minimal 276-520 region sufficient for spatial regulation in the *XCR1α* 3'UTR is presented in Figure 3.36. I have shown that three key factors are necessary for spatial regulation: a spatial regulator binding TATTAC at position 360 in the 3'UTR, an RNA-stabilising factor and an additional spatial regulator between nucleotides 450-520. These results are in some aspects complementary to another model of *XCR1α* 3'UTR regulation published whilst this work was in progress (Figure 3.37) (Zhang et al., 2009).

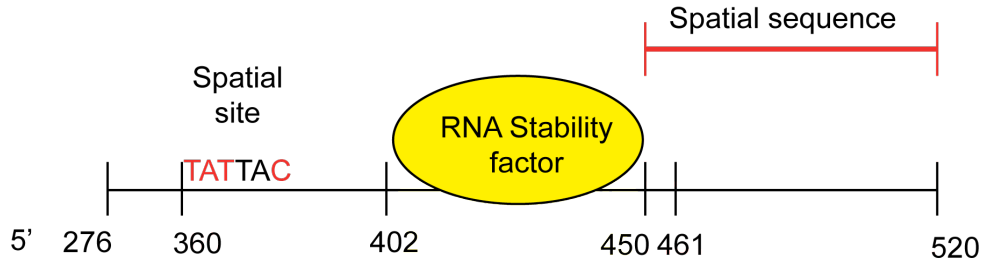


Figure 3.36 Spatial regulatory sites within the minimal 276-520 nt region of the *XCR1α* 3'UTR sufficient for spatial regulation
The mutated nucleotides necessary for spatial regulation beginning at nucleotide position 360 are shown in red. An RNA-stabilising factor binds between nt 402-450, possibly an RNA-binding protein. A further sequence located between nucleotides 450-520 is necessary for spatial regulation, shown in red brackets.

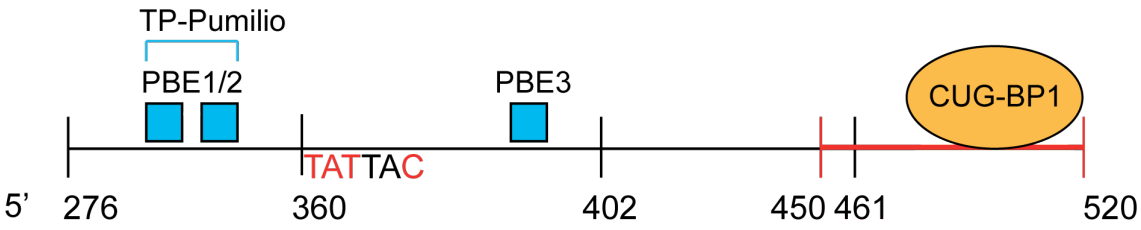


Figure 3.37 Zhang *et al.* show that Pumilio core sequences and CUG-BP1 sites are necessary but not sufficient for spatial regulation
Pumilio binding elements (PBE) with a core sequence of TGTA (shown in blue) are located near the TATTAC spatial site I identified but do not overlap. The target protector oligo I used (TP-Pumilio shown by a blue bracket) which had no effect on spatial regulation was located over the PBE1/2 core TGTA sequences. The RNA-binding protein CUG-BP1 was also shown to be necessary but not sufficient by Zhang *et al.*, and is located in the spatial sequence I identified (450-520).

Zhang *et al.* show that the same region of the *XCR1α* 3'UTR is sufficient for spatial regulation, nucleotides 286-637, which I have refined to 276-520 nt. Zhang *et al.* show that mutating the core TGTA region of three Pumilio binding elements (PBEs) and the binding site for the negative translational regulator CUG-BP1 protein in the context of 286-637 nt partially relieves repression of vegetal translation of *XCR1* mRNA (Zhang et al., 2009). They conclude that the Pumilio and CUG-BP1 sites are necessary but not sufficient for spatial regulation. The four nucleotides I identified as necessary for spatial regulation (TATTAC, Figure 3.36) do not overlap with the core TGTA Pumilio regions mutated by Zhang *et al.* Moreover, blocking Pumilio binding with a target protector over the PBE1/2 site has no effect on spatial regulation (Figure 3.29). Since TP-Pumilio has no effect and mutating the TATTAC site at 360 nt completely reverts spatial regulation in a GFP-3'UTR reporter mRNA assay, which is a visual readout of spatial regulation, rather than an indirect luciferase-3'UTR assay requiring dissection of the embryo used by Zhang *et al.*, I conclude that an unknown factor binding the TATTAC site at 360nt is necessary for spatial regulation. Mutating Pumilio core elements near this site may alter the secondary structure of the mRNA, contributing to the partial reversion of spatial regulation indirectly observed by Zhang *et al.*

Interestingly, Zhang *et al.* also show that a CUG-BP1 RNA binding protein site is also necessary for spatial regulation. The CUG-BP1 site is within the 450-520 nt region

I identified as the second sequence necessary for spatial regulation and could therefore be a necessary factor for spatial regulation. This could be studied by knocking down CUG-BP1 and assaying for loss of spatial regulation. Although Zhang *et al.* do identify the CUG-BP1 site as necessary for spatial regulation, no evidence is presented that the protein itself plays any role in spatial regulation. Similarly, Pumilio knockdown data would strengthen the model presented by Zhang *et al.* In short, the mapping data from both studies is complementary, although the regulator binding the TATTAC site at 360 nt has yet to be conclusively identified.

3.3.4 Potential spatial regulators of XCR1

Potential regulators for binding the RNA-stabilising sequence between 402-450nt in the *XCR1α* 3'UTR included the poly(U) binding protein (Pub1) and the RNA-binding protein Hur. The mutational analysis in Figure 3.28 showed that mutating the overlapping predicted binding sites from a T/U rich sequence to 5'TTTTATATGAACTTAATCAATAC had no effect on spatial regulation or RNA stability assessed by GFP levels. Since both Pub1 and Hur bind AU-rich elements (AREs), mutating bases from T to A may not have been the optimal strategy to test the effect of Pub1, Hur or any ARE-binding protein on spatial regulation and/or RNA stability. Mutations from T to G/C should be tested before the role of ARE binding proteins, which recruit the exosome to degrade target RNA (Chen and Shyu, 1995), on RNA stability in the context of the *XCR1α* 3'UTR is ruled out.

A scan of the full length *XCR1α* 3'UTR sequence against the database of RNA protein binding specificities, (RBPDB) (Cook et al., 2010) yielded 112 hits with a score greater than 80 % of the maximum possible score for the protein. RNA binding protein hits within the 276-520 nt region are presented in Table 3.5. There are 7 Hur/ELAV motif sites within 402-461 RNA-stabilising sequence and a site for Poly(A) binding protein at 461nt which are strong candidates for stabilising mRNA. Notably, the Mut-107 site at 285-290 nt disrupts a motif for the pre-mRNA splicing regulator MBNL1, which may explain why Mut-107 caused a decrease in GFP expression to undetectable levels in the context of the 276-520 3'UTR. Intriguingly, the RNA-binding protein QKI, which can regulate translation as well as RNA stability and localisation (Biedermann et al., 2010), has a motif at 352-361 nt which overlaps the necessary spatial four

nucleotides at 360 nt and may therefore play a role in spatial regulation. An RNA pull-down experiment in the context of wild-type and Mut-1810 full length *XCR1α* 3'UTR and subsequent mass spectrometry or bandshift analysis would reveal whether this site was a critical RNA-binding protein sequence for spatial regulation.

There are both *Xenopus laevis* and *Xenopus tropicalis* homologues of QKI, *qki-a, qki-b* and *qki* respectively (Bowes et al., 2010). Mice mutant for *QKI* display defective myelination of neurons and 'quaking' tremors (Larocque et al., 2002). Further analysis revealed that in *QKI* mutant mice, *myelin basic protein (MBP)* mRNA is retained in the nucleus instead of exported to the cytoplasm (Larocque et al., 2002). As well as regulating mRNA localisation, QKI is also known to stabilise *p27(Kip1)* mRNA to increase levels of the protein (Larocque et al., 2005). The most relevant reported function of QKI proteins for the mechanism of XCR1 spatial regulation is probably that of the *C. elegans* QKI protein, GLD-1, which acts as a translational repressor for cyclin E (Biedermann et al., 2009), p53 (Schumacher et al., 2005) and multiple early development mRNA targets including one target where spatial repression of translation is observed (RME-2) (Lee and Schedl, 2001). Intriguingly, in mice, QKI-6 has been found to interact with Ago-2, which suggests that QKI proteins may interact with the microRNA-targeting pathway (Wang et al., 2010b). In short, the possibility that *Xenopus laevis* QKI proteins are responsible for spatial regulation of XCR1 could be tested by loss of function studies with morpholinos and assaying for spatial regulation.

Relative Score	RBP Name	Start	End	Matching sequence	Comment
100%	ELAVL2	399	407	UUUUAUUUU	Also known as Hur
100%	ELAVL2	399	407	UUUUAUUUU	
80%	ELAVL2	400	413	UUUAUUUUUUUUUA	
88%	ELAVL2	409	417	UUUUAAUUU	
100%	A2BP1	482	487	UGCAUG	Regulates alternative splicing
100%	A2BP1	288	293	UGCAUG	
100%	a2bp1	483	487	GCAUG	Zebrafish ATBP1 homologue
100%	a2bp1	289	293	GCAUG	
94%	PABPC1	461	467	ACAAAUA	Poly(A)binding protein involved in translation initiation regulation via PAIP1/2
88%	RBMY1A1	392	396	CACAA	
100%	Pum2	502	505	UGUA	
100%	Pum2	343	346	UGUA	
100%	Pum2	372	375	UGUA	
100%	Pum2	335	338	UGUA	
100%	MBNL1	278	281	UGCU	
100%	MBNL1	285	288	UGCU	pre-mRNA splicing
100%	EIF4B	303	306	GGAA	
100%	EIF4B	451	454	GGAA	
100%	KHSRP	470	473	GUCC	Splicing regulation
100%	KHSRP	511	514	GUCC	
100%	RBMX	324	327	CCAG	
81%	QKI	352	361	UAUUAACUUA	pre-mRNA splicing, mRNA export, mRNA stability and protein translation
83%	YTHDC1	384	389	GCAUCC	
80%	YTHDC1	474	479	UACUAC	Splice site selection
100%	SFRS1	302	305	AGGA	
100%	SFRS1	450	453	AGGA	Alternative splicing
88%	SFRS13A	434	440	AGAGAAC	Repressor of pre-mRNA splicing
88%	SFRS13A	432	438	AAAGAGA	
94%	KHDRBS3	422	427	AUUAAU	Alternative splicing regulator
100%	ELAVL1	507	510	GUUU	

100%	ELAVL1	499	502	GUUU	
93%	KHDRBS3	410	415	UUUAAU	
86%	ELAVL1	403	406	AUUU	
86%	ELAVL1	414	417	AUUU	
86%	ELAVL1	456	459	AUUU	
86%	ELAVL1	348	351	AUUU	
82%	KHDRBS3	343	348	UGUAAA	
81%	KHDRBS3	377	382	AUUAAC	
81%	KHDRBS3	353	358	AUUAAC	

Table 3.5 RNA binding protein motif matches in the 276-520 region of the *XCR1α* 3'UTR as predicted by the RBPDB database with an 80 % relative score threshold. Proteins referred to in the text are highlighted in yellow and functions as described in the Uniprot database are annotated.

In order to expand the list of possible spatial microRNA regulators from those tested so far, 10 nucelotides of 3'UTR sequence was added either side of the four nucleotide spatial sequence at 360 nt and scanned for seed matches in all miRbase microRNAs. This was to include microRNAs which were not present in the *X. tropicalis* small RNA-seq libraries and microRNAs where the Mut-1810 mutations would disrupt microRNA binding, but not necessarily in the entire seed sequence. As shown in Table 3.6, miR-187 was the microRNA with the most number of seed matches across species including *X. tropicalis*, humans and other mammalian species that could be disrupted by mutating the 5'TATTAG spatial site in the *XCR1α* 3'UTR. Mouse miR-187 was sequenced with 12 reads in the *X. tropicalis* stage 10 small RNA library. Other microRNAs were isolated examples with unique seeds (e.g. osa-miR-440) or microRNA families in a single plant species such as the ptc-miR-478 family example. miR-187 is therefore a candidate spatial regulator because its binding to the *XCR1α* 3'UTR is predicted to be disrupted by the Mut-1810 manipulation and could be tested by miR-187 loss of function experiments.

Other candidate microRNAs regulating the *XCR1α* 3'UTR can be added by taking microRNAs that are expressed at stage 10 in the small RNA-seq (relaxing the criteria of presence in both the stage 8 and stage 10 libraries) and looking for exact seed matches to the full length 3'UTR sequence (Table 3.7). There are 12 microRNAs with exact seed matches in the 276-520 minimal region of the *XCR1α* 3'UTR that have not yet been tested, including mmu-miR-187. These candidate spatial regulators could be tested by

mutating the predicted miRNA seed binding site and loss of function studies with microRNA morpholinos.

In summary, there are multiple candidates including RNA binding proteins and microRNAs that could mediate spatial regulation of XCR1 through the *XCR1α* 3'UTR that remain to be experimentally investigated. I have shown that spatial regulation of XCR1 is a result of repression of translation in the vegetal pole at stage 10, partly requiring Dicer, and two spatial regulatory sequence elements and an RNA stabilising element in the *XCR1α* 3'UTR are necessary. Uncovering the mechanism of spatial regulation of XCR1 highlighted the lack of knowledge about microRNA expression in early development, which led me to perform genome-wide small RNA sequencing of early development small RNA libraries, as discussed in Chapter 4.

miRNA seed sequence	**Seed match proximity to 5'tattag spatial site in 3'UTR**	**Number of miRNAs with that seed**	**miRNA**
5'CGTGTC	5'TTTATTAACTtatta[gACACG]TTGTA	20	miR-187 in many species including: xtr-miR-187, hsa-miR-187 and smo-miR-536
5'GTGTCT	5'TTTATTAACTtatt[agACAC]GTTGTA	1	osa-miR-440
5'AACGTG	5'TTTATTAACTtattagA[CACGTT]GTA	13	ptc-miR-478 family

Table 3.6 miR-187 is the best candidate for a microRNA a seed match proximate to the spatial four nucleotides sequence at 360nt
The spatial site 'tattag' is shown in lower case and the square brackets indicate the position of the microRNA seed match in the *XCR1α* 3' UTR sequence.

***X. tropicalis* st 10 miRNA sequence**	**Read number**	**Start of seed alignment in *XCR1α* 3'UTR (nt)**
cel-miR-72	113	717
cel-miR-73	52	717
cel-miR-74	8	717
hsa-miR-26a	28	13
hsa-miR-101	49	884, 936
mmu-miR-101a	12	884, 936
mmu-miR-124	12	160
mmu-miR-125a-5p	12	843
mmu-miR-125b-5p	12	843
mmu-miR-126-5p	12	849
mmu-miR-130a	12	61
mmu-miR-135a	12	542
mmu-miR-136	12	695
mmu-miR-144	12	884, 936
mmu-miR-152	12	248
mmu-miR-153	12	142, 595
mmu-miR-154	12	707
mmu-miR-10b	12	898
mmu-miR-181a	12	658, 798
mmu-miR-186	12	775
mmu-miR-187	12	370
mmu-miR-193	12	23
mmu-miR-199a-5p	12	249
mmu-miR-199a-3p	12	271
cel-miR-237	33	843
dme-miR-263a	15	381, 929
dme-miR-279	15	313
dme-miR-124	233	160
mmu-miR-106a	12	63, 603
mmu-miR-106b	12	63, 603
mmu-miR-130b	12	61
dme-miR-125	233	843
dme-miR-31a	41	717
dme-miR-312	76	918
mmu-miR-148a	12	248
mmu-miR-103	12	290
mmu-miR-148b	12	248
mmu-miR-135b	12	542
mmu-miR-101b	12	884, 936
mmu-miR-107	12	290
mmu-miR-10a	12	898
mmu-miR-17	12	63, 603

mmu-miR-199b	12	271
mmu-miR-181b	12	658, 798
mmu-miR-181c	12	658, 798
cel-miR-35	31	329
hsa-miR-363	8	918
gga-miR-30a-3p	428	135
gga-miR-101	65	885, 937
dre-miR-10a	445	898
xla-miR-427	92	62
dre-miR-10c	1	898
bta-miR-29a	30	633
xtr-miR-10a	521	898
xtr-miR-10b	342	898
xtr-miR-17-5p	1	63, 603
xtr-miR-26	22	13
xtr-miR-30a-3p	3	135
xtr-miR-30b	13	231
xtr-miR-107	11	290
xtr-miR-130a	14	61
xtr-miR-130b	12	61
xtr-miR-202	2	719
xtr-miR-181a	1	658, 798
xtr-miR-101a	189	884, 936
xtr-miR-148a	189	248
xtr-miR-25	12	918
xtr-miR-31	3	717
xtr-miR-20b	1	63, 603
bta-miR-107	19	290
bta-miR-140	99	494, 642
sme-miR-31b	20	717
mmu-miR-181d	12	658, 798
mmu-miR-193b	12	23
dme-miR-193	233	23
dme-miR-1002	233	322
dme-miR-137	233	623
dme-miR-1001	233	325
dme-miR-1003	233	558
dme-miR-1004	233	558
dme-miR-1005	233	809
dme-miR-1013	233	403, 777
xtr-miR-31	3	280, 716
mmu-miR-466i	35	210
mmu-miR-1-2-as	12	508, 888
mmu-miR-1186	12	247
mmu-miR-1197	12	474, 515
oan-miR-31	5	717

cfa-miR-31	55	717
cfa-miR-101	12	884, 936
cfa-miR-30e	1038	135
cfa-miR-140	51	641
cfa-miR-363	32	918
hsa-miR-182	12	249
mmu-miR-1905	12	761
mmu-miR-1906	12	290
mmu-miR-1904	12	145
mmu-miR-1930	12	696
mmu-miR-1933-3p	12	516
mmu-miR-1935	12	20
mmu-miR-1944	12	823
mmu-miR-1950	12	143
mmu-miR-1959	12	578
mmu-miR-1960	12	248
mmu-miR-1971	12	723
mmu-miR-1981	11	723
mmu-miR-1982.2	12	899
spu-miR-10	6	898
bfl-miR-2076	40	918
dpu-miR-31	10	717

Table 3.7 110 stage 10 microRNAs that bind the full length *XCR1α* 3'UTR with an exact seed match

12 miRNAs with seed matches in the spatial 276-520 spatial region of the 3'UTR not yet studied are highlighted in yellow. MicroRNAs are those passing the filters of 22-23nt in length, with zero mismatches to a miRbase version 14 microRNA sequenced at stage 10 with an exact seed match in the *XCR1α* 3'UTR.

Chapter 4. Genome-wide small RNA sequencing of the early *Xenopus* embryo reveals dynamic and localised expression of small RNAs

4.1 Introduction

The aim of the work described in this chapter was to sequence all small RNAs in the *Xenopus tropicalis* embryo at three stages in early development: blastula (stage 8), gastrula (stage 10), neurula (stage 18) and in animal and vegetal hemispheres of the gastrula stage embryo. This project emerged from working on the mechanism of spatial regulation of *Xenopus Cripto-1* in early development. Very few *Xenopus* microRNAs were known in early development, as most microRNAs had been cloned in adult tissues (Watanabe et al., 2005). To fill this void, I sequenced all small RNAs present at three stages of *X. tropicalis* early development to characterise new *Xenopus* microRNAs and obtain dynamic and spatial information on microRNA expression and expression of other small RNAs during early development.

I prepared small RNA libraries size selected for 18-30 nucleotides in length as described in Chapter 2 and Figure 4.1 which were sequenced using the Illumina next-generation sequencing platform. The small RNA libraries were designed to contain microRNAs (~22 nt), endogenous siRNAs (21-24 nt) and any Piwi-interacting RNAs (piRNAs, 25-30 nt) plus any other unknown small RNA species. *Xenopus tropicalis* rather than *Xenopus laevis* libraries were prepared due to the availability of the *X. tropicalis* reference genome (Hellsten et al., 2010). Characterisation and experimental validation of the small RNA-seq dataset is presented in this chapter. Bioinformatic analysis of the raw dataset was performed by Stuart Horswell in the Cancer Research UK Bioinformatics and Biostatistics Group and I designed the downstream analysis of the dataset.

4.2 Results

4.2.1 Quality of the small RNA libraries

Firstly, the quality of the small RNA libraries was assessed. Following ligation of Illumina adapter sequences onto the small RNAs and reverse-transcription PCR

amplification (Figure 4.1), the final stage of small RNA library preparation showed strong bands at ~100 nt in each of the libraries (Figure 4.2). Adapter dimer bands were avoided, and the small RNA cDNA libraries were gel purified and quantified for Illumina sequencing. Libraries were submitted at a 10 nM concentration for single-end small RNA sequencing using the Illumina platform on a Genome Analyzer IIx at the Cancer Research UK Cambridge Research Institute. A technical repeat of the Illumina sequencing using the same libraries was performed approximately one year later on a Genome Analyzer IIx at the Cancer Research UK London Research Institute. All analysis presented in this chapter refers to the results from the first sequencing run, except where the datasets are directly compared.

The FastQ files from the initial sequencing run were analysed. Over 14 million small RNA raw reads were obtained per library. These were filtered to retain reads that contained at least six contiguous nucleotides of Illumina adapter sequence and had at least 15 nucleotides of sequence remaining when the adapter sequence was removed. Approximately 70 % of raw reads were retained after adapter trimming in each library (Table 4.1). Trimmed reads were then aligned to version 4.1 of the *X. tropicalis* genome with zero mismatches permitted. Between 48.8 % and 59.3 % of the raw reads passed this strict alignment criteria (Table 4.2). Less than 0.1 % of untrimmed reads in each library aligned with no mismatches to the genome. This is an estimate of the false detection rate, as untrimmed reads contain fragments of the unique Illumina adapter sequences at the 3' end that do not align anywhere in the genome.

Small RNA reads can be compressed into tags, which are unique sequences. For example, a small RNA library could be comprised of millions of reads which are many copies of the same sequence and would therefore collapse into one tag. In the early *Xenopus tropicalis* embryo, the number of tags that align perfectly to the genome ranges between 387,173 at stage 10 to 718,228 in the stage 10 vegetal pole library. Another feature of the dataset obvious at this stage in the analysis is high multiplicity: when a read aligns to more than one place in the genome. For example, as shown in Table 4.2, at stage 8, over 670,000 tags align to more than 26 million places in the genome. This could be due to alignment to repetitive regions in the genome or multiple loci for small RNA genes.

The quality of the data set was enhanced by biological replicates within the libraries, as there are a total of three stage 10 samples (stage 10 whole embryo, stage 10 animal pole and stage 10 vegetal pole), which should overlap in small RNA content (see section 4.2.2.1). A technical repeat of the small RNA sequencing performed one year later at the Cancer Research UK London Research Institute yielded greater sequencing depth, indicated by the increase in raw read numbers from around 14 million to over 20 million per library (Table 4.3). The percentage of raw reads aligning perfectly to the genome was similar to the first run at 45-57 % and the number of aligned tags followed the same pattern of tags decreasing between stage 8 and 10, then increasing at stage 18 with more vegetal than animal tags at stage 10. An increase in tags resulted from the re-sequencing for the stage 8, stage 18 and stage 10 vegetal libraries. The additional microRNAs tags identified by re-sequencing the libraries are presented in Table 4.7.

Finally, an immediate picture of the quality of the libraries is given by the length distribution of small RNA species sequenced, shown for trimmed, aligned reads and tags in the first sequencing run at the CRI, and for the technical repeat performed at the LRI (Figure 4.3). Firstly, the accuracy of the size selection of small RNAs of length 18-30 nt can be assessed. The reads are trimmed of adapter sequence so some of the sequences in the length distribution fall between 13-18 nt. Secondly, an impression of heterogeneity within the library is revealed. Peaks in the length distribution can suggest the presence of different classes of small RNAs in the libraries. Thirdly, the libraries can be compared to one another. Figure 4.3 shows that the raw read and tag length distributions are similar for all the libraries, with the stage 10 library containing the fewest tags.

The length distribution of reads in the first sequencing run (CRI reads) shows a major peak at 28 nucleotides in all libraries, and a lesser peak at 23 nt in all except the stage 10 animal library. These sizes correspond to piRNAs and microRNAs respectively (Armisen et al., 2009). The 28 nt peak does not collapse when raw reads are compressed into tags, indicating a complex population of small RNAs characteristic of piRNAs. The 23 nt peak does collapse, which is a feature of microRNA reads, as many identical microRNA sequences compress into far fewer tags (Armisen et al., 2009). When the initial sequencing run and technical repeat length distributions are compared, it is obvious that the read length profiles share very similar features, except that the 23 nt

peak is not apparent in the technical repeat. The tag profiles for both sequencing runs are similar, with the technical repeat showing lower abundance of 28 nt tags compared to the initial sequencing run. In short, the length distributions from both runs of small RNA sequencing are highly reproducible.

In summary, the high quality of the small RNA libraries was established by alignment of the small RNA-seq reads to the *X. tropicalis* genome and by the length distribution of small RNAs within the library. Re-sequencing of the libraries in a technical repeat reiterated and added to the findings of the initial study.

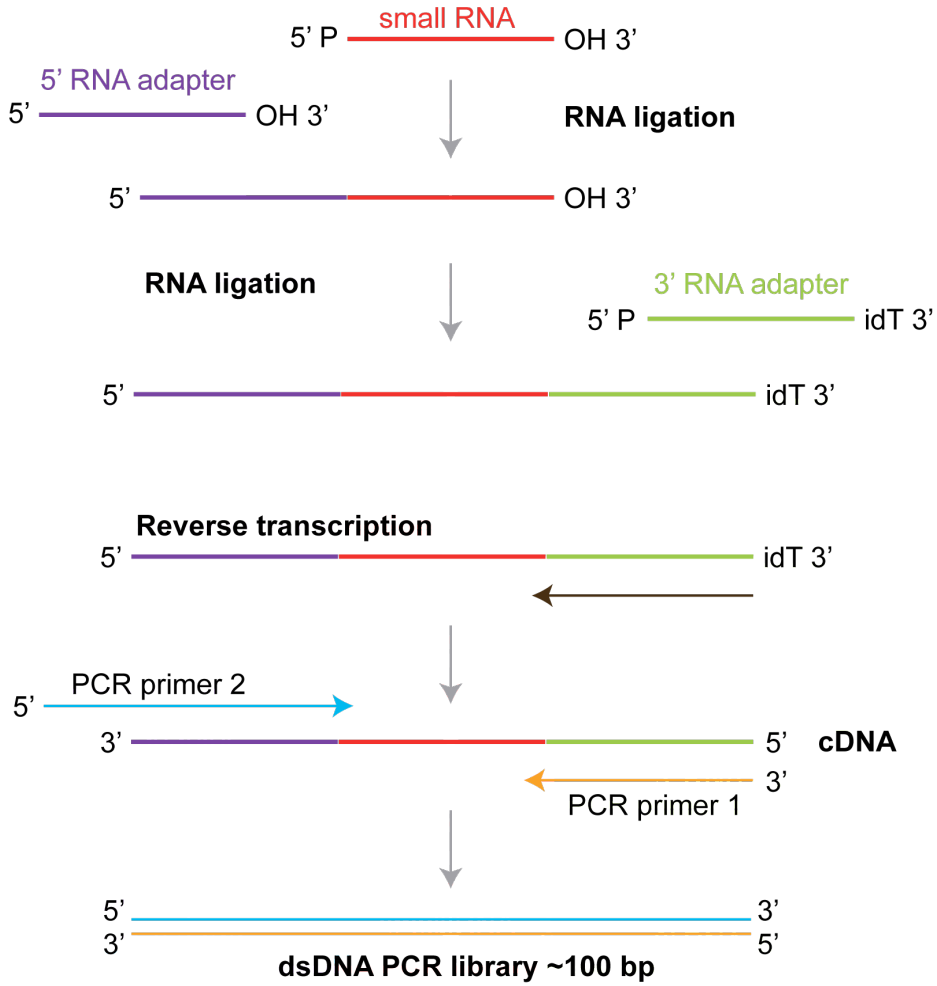


Figure 4.1 Overview of Illumina small RNA library construction

Schematic diagram of Illumina small RNA library construction. The protocol and primer sequences are described in section 2.6. Briefly, small RNAs of 18-30 nt length are purified and ligated onto 5' and 3' Illumina RNA adapters by an RNA ligation reaction. The adapter sequence is needed to stick the template DNA to the flow cell in the Illumina Genome Analyser IIx sequencing machine. The 3' RNA adapter contains a modified base, idT. The ligated small RNAs are then reverse transcribed and the resulting cDNA is used as the template for a large scale PCR reaction using Illumina PCR primers 1 and 2. The final PCR product is about 100 bp (see Figure 4.2). The dsDNA PCR library of small RNAs is purified and sent for Illumina sequencing.

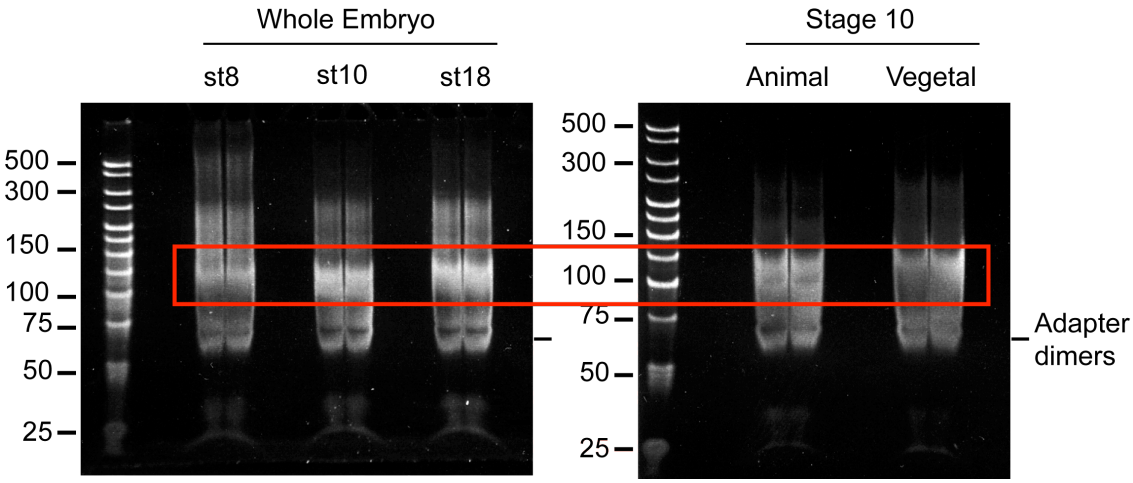


Figure 4.2 Final quality control stage of small RNA library preparation

After both Illumina adapters were ligated on and cDNA libraries were PCR-amplified as described in Figure 4.1, dsDNA PCR small RNA libraries were gel purified. The PCR products are run on a 6 % native polyacrylamide gel and stained with ethidium bromide. The ~100 bp band (red box) is excised, avoiding the adapter dimers at ~ 70 bp. The integrity and successful preparation of the library was confirmed at this stage before final quantification and Illumina sequencing.

Small RNA library	Number of Raw Reads	Number of Reads after Adapter Trimming	Trimmed Reads as % of Raw Reads
Stage 8	15,824,233	11,553,093	73.0
Stage 10	15,338,864	11,634,636	75.9
Stage 18	14,739,870	10,463,257	71.0
Stage 10 animal	14,297,841	10,277,485	71.9
Stage 10 vegetal	15,171,802	10,350,900	68.2

Table 4.1 Approximately 70 % of raw reads are retained after adapter trimming

Raw reads were filtered by retaining reads with at least six contiguous nucleotides of Illumina adapter sequence at the 3' end with no less than 15 nt of sequence remaining. Adapter sequences were removed from raw reads to give trimmed reads.

Small RNA library	Number of Reads aligning to the genome (0 mismatches)	Aligned Reads as % of Raw Reads	Number of Aligned Tags	Multiple Alignments
Stage 8	9,125,752	57.7	672,032	26,419,988
Stage 10	9,092,783	59.3	387,173	16,331,788
Stage 18	8,067,818	54.7	566,646	26,770,113
Stage 10 animal	8,182,367	57.2	551,798	25,413,967
Stage 10 vegetal	7,399,178	48.8	718,228	33,261,573

Table 4.2 Alignment of trimmed reads to the *Xenopus tropicalis* genome

Trimmed reads were aligned to the *X. tropicalis* genome with zero mismatches. Approximately 50-60 % of the raw reads aligned to at least one place in the genome with no mismatches after adapter trimming. Tags often align to multiple places in the genome. The multiple alignments column indicates the total number of alignments for all tags.

Small RNA library	Number of Raw Reads	Trimmed Reads as % of Raw Reads	Aligned Reads as % of Raw Reads	Number of Aligned Tags
Stage 8	23,119,446	67.1	54.7	696,569
Stage 10	21,565,804	71.5	57.1	358,909
Stage 18	27,893,758	61.5	48.9	599,334
Stage 10 animal	20,704,206	68.3	56.0	528,987
Stage 10 vegetal	26,455,794	60.4	44.8	834,281

Table 4.3 A technical repeat of library sequencing shows improved sequencing depth

Raw read numbers improved from approximately15 million to over 20 million raw reads on re-sequencing of the small RNA libraries. The alignment percentages and are comparable to the initial sequencing run (Table 4.2). The number of aligned tags increases in the stage 8, stage 18, and stage 10 animal libraries and decreases in the stage 10 and stage 10 animal libraries on re-sequencing, although the number of tags is broadly similar to the initial run.

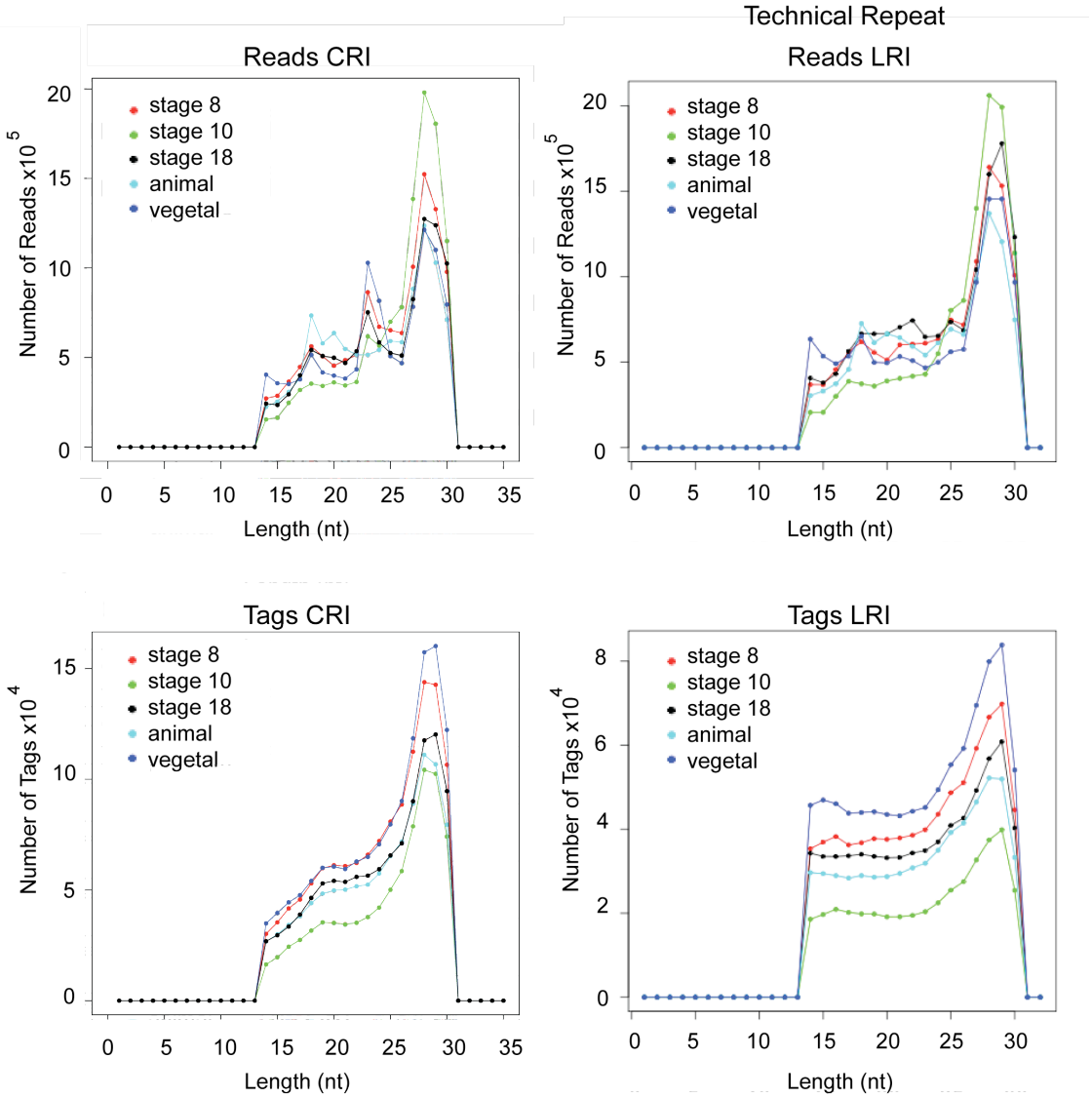


Figure 4.3 Length distributions of trimmed, aligned Reads and Tags in the small RNA libraries

Reads can be compressed into unique sequences known as 'tags'. Read and tag number is plotted against read length for the stage 8, stage 10, stage 18, stage 10 animal and stage 10 vegetal libraries for the first sequencing run (CRI) and the technical repeat sequencing run (LRI). The reads and tags are trimmed to remove adapter sequences and align to the genome with zero mismatches. The read profiles for all small RNA libraries in the LRI dataset contain peaks at ~23 nt and 28 nt, whereas the tag profile shows one major peak at 28 nt.

4.2.2 Overview of the *X. tropicalis* early development small RNAome

In order to annotate the small RNAs sequenced in the five libraries, trimmed reads that aligned to the *X. tropicalis* genome with zero mismatches were aligned to non-coding RNA databases. Firstly, the mature miRNA sequences contained in the miRbase miRNA database for all species (Griffiths-Jones et al., 2006) were compressed into unique sequences, retaining a *Xenopus* tag when a microRNA was also conserved in other species. This left 5,213 miRbase miRNA tags of which 156 were *Xenopus tropicalis* miRNAs. Trimmed reads were then aligned to these miRbase tags and normalised to allow comparison across the small RNA libraries by dividing the number of aligned reads by the total number of reads which aligned to the genome in each library and multiplying by $1x10^7$. The number of reads aligning to a miRbase microRNA in each library and the number of microRNA tags in each library is shown in Figure 4.4A.

Both microRNA reads and tags were highest in the stage 18 library, suggesting that microRNAs increase in abundance and sequence diversity between gastrulation and late neurula stage in *X. tropicalis*. A maximum of 851 microRNA tags was obtained in the stage 18 library in this initial filtering, of which 75 were already known in *X. tropicalis*. Therefore, 48 % of the known *X. tropicalis* miRNAs were sequenced and are present in early development. More microRNA reads and tags were present in the stage 10 animal pole library than in the stage 10 vegetal pole library, which will be explored further in Section 4.2.2.1. MicroRNA-star and hairpin sequences were also downloaded from miRbase as were microRNAs in the RNAdb database (Pang et al., 2005) and aligned to trimmed reads. In total, all microRNA-related sequences comprised less than 1-2 % of the small RNAome in all the libraries. MicroRNAs were maximally 1.56 % of the trimmed reads at stage 18 (Table 4.4).

Next, the reads annotated as mature microRNAs, miRNA-star or hairpin sequences were removed and the unannotated trimmed reads left were compared to the RNA family database, Rfam, a repository of non-coding RNAs (Griffiths-Jones et al., 2003). Reads aligning to Rfam RNA sequences were the most abundant of the annotated sequences in the *X. tropicalis* small RNAome, seen by the high normalised read

numbers in the libraries (Figure 4.4C), and the percentage of trimmed reads aligning which ranged from 1.1 % at stage 8 to 2.2 % in the stage 10 animal pole library (Table 4.4). When the categories of non-coding RNAs hit by the trimmed reads were analysed (Figure 4.5), the profiles were consistent across the five small RNA libraries and the most frequently hit Rfam tags were ribosomal RNAs (38-43 %), tRNA (10-14 %) and small nucleolar RNA (snRNA, 10-13 %). Since the small RNAs cloned were 18-30 nt, these are alignments to part of the Rfam tag sequence.

The remaining unannotated reads were then aligned to Piwi-interacting RNAs in RNAdb. piRNA reads were highest in the stage 10 animal pole library, and increased from stage 10 to stage 18 (Figure 4.4B). Interestingly, piRNA tags decreased between stage 8 and stage 10, and the number of piRNA unique sequences was higher in the stage 10 vegetal pole library compared to the animal pole, despite the lower piRNA read numbers. This suggests that some piRNAs show stage-specific expression and the vegetal pole contains a more complex population of piRNAs at lower abundances than the animal pole. Piwi-interacting RNAs were a maximum of 0.88 % of the small RNAome in the stage 10 animal library.

Tags from the other non-coding RNA datasets in RNAdb were sequentially extracted from the Hinv, Fantom3, Evofold, antisense non-coding RNA pipeline and literature-curated datasets (Pang et al., 2005). Fantom3 tags were hit by up to 0.49 % of trimmed reads, whereas the sum of the other database tag hits was minimal, at less than 1.1 $x10^{-3}$ %. Significantly, between 95 % and 98 % of small RNA trimmed reads, which align to the genome, remained unannotated following comparison to known non-coding RNA databases. This result was confirmed by analysis of the technical repeat sequencing run, where between 91% and 96 % of the reads that aligned to the genome were unannotated (Table 4.5). The percentage of microRNAs and piRNA reads was very similar in the technical repeat, but more reads aligned to Rfam and Fantom3 than in the first sequencing run, reducing the number of unannotated reads.

In summary, most of the *X. tropicalis* small RNAome in early development is as yet uncharacterised, and the question of what these small RNAs are, and their role in early development will be addressed in Chapter 5.

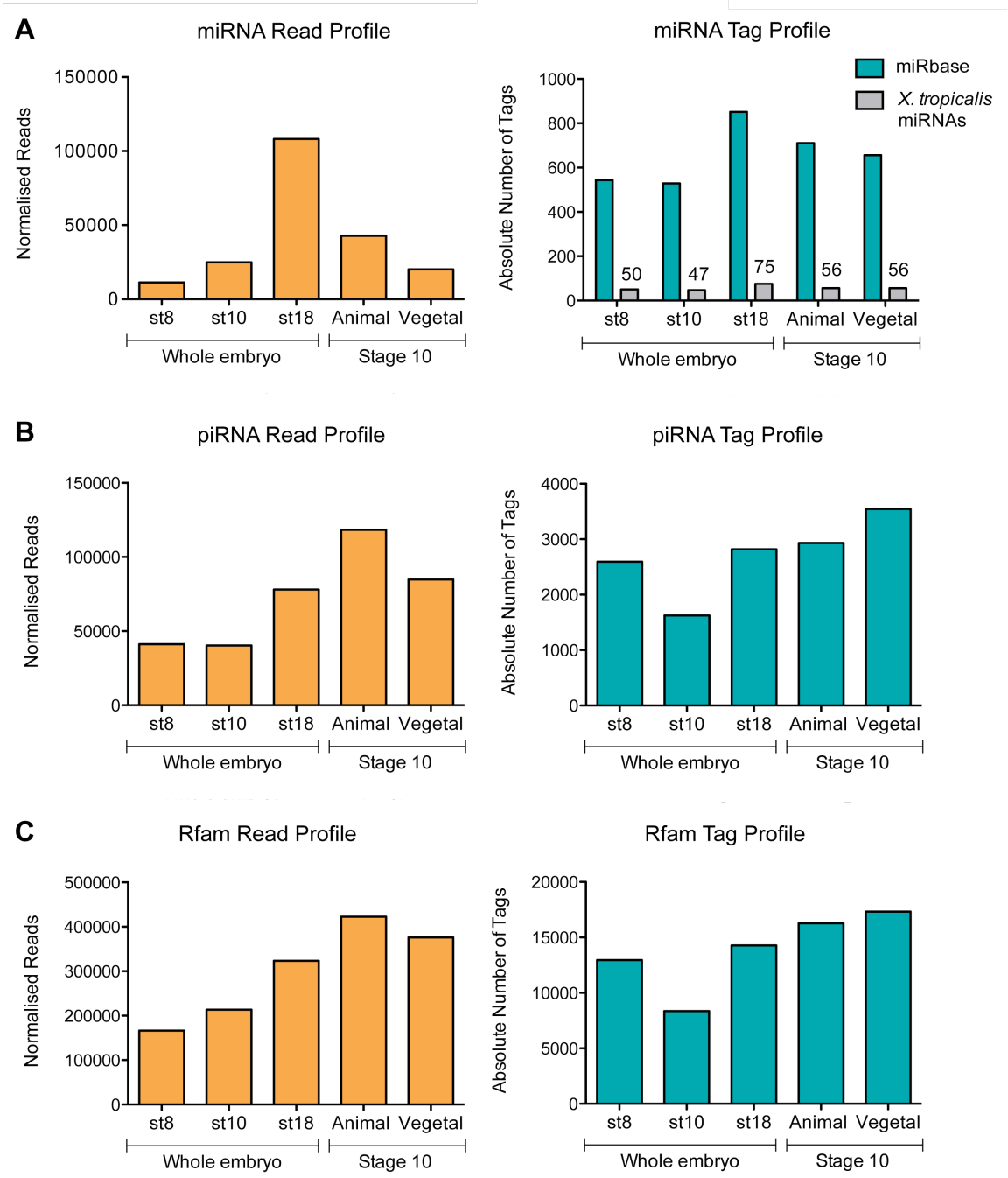


Figure 4.4 Read and Tag profiles across the libraries for microRNAs, piRNAs and Rfam small RNAs.

A. Trimmed reads were aligned to microRNAs in miRbase version 14 (5,213 unique sequences) and normalised to allow comparison across the libraries by dividing the number of miRNA reads by the total number of reads aligning perfectly to the genome and multiplying by $1x10^7$. The absolute number of miRbase miRNA tags sequenced is shown, with known *X. tropicalis* miRNA tags in grey (156 possible sequences). **B.** Reads were aligned to the piRNAs in RNAdb. **C.** Reads were aligned to non-coding RNAs in the RNA families (Rfam) database.

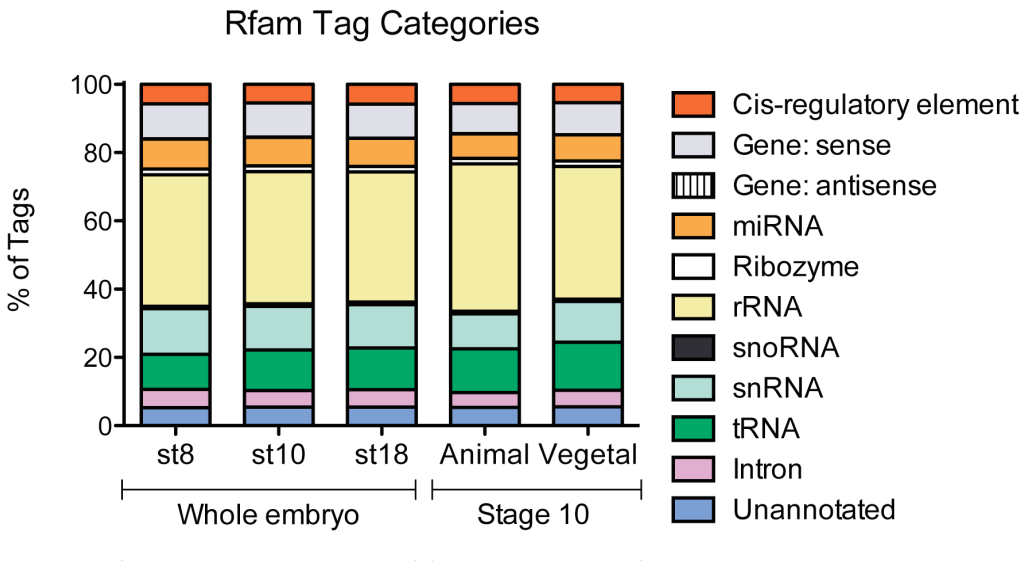


Figure 4.5 Percentage breakdown of Tags aligning to Rfam database small RNAs

Tags matching to Rfam RNA sequences in the stage 8, stage 10, stage 18 and stage 10 animal and vegetal libraries were analysed for which Rfam category they corresponded to. The 'cis-regulatory element' Rfam category contains RNA sequences with a known cis-regulatory function, such as stem-loops in 3' UTRs that interact with proteins. The 'gene' category contains non-coding RNA genes and the alignment to tags is broken down into a 'sense' and 'antisense' alignment. MicroRNAs present in the Rfam database do not overlap with miRbase sequences as the latter were extracted prior to Rfam alignments. The 'ribozyme' category contains RNA enzyme sequences. The 'rRNA' category contains ribosomal RNA sequences. 'snoRNA' refers to small nucleolar RNA sequences, whereas the 'snRNA' category contains small nuclear RNA sequences. The 'tRNA' category contains transfer RNA sequences. The 'intron' Rfam category contains RNA sequences originating from introns such as self-splicing catalytic introns. The 'unannotated' Rfam category contains sequences that have not been assigned to any RNA family.

Small RNA Library	% miRNA reads total	% piRNA reads	% Rfam reads	% Fantom3 reads	% Hinv, Evofold, Antisense ncRNA pipeline and literature data set reads	% Un-annotated genomic reads
Stage 8	0.17	0.37	1.10	0.21	3.8×10^{-4}	98.1
Stage 10	0.32	0.31	1.19	0.15	5.0×10^{-4}	98.0
Stage 18	1.56	0.56	1.17	0.26	3.3×10^{-4}	96.5
Animal	0.67	0.88	2.20	0.49	1.1×10^{-3}	95.8
Vegetal	0.33	0.68	1.41	0.32	8.2×10^{-4}	97.3

Table 4.4 *X. tropicalis* early development small RNAome annotation

Summary of the percentage of trimmed reads that align with zero mismatches to miRNA sequences (mature miRNAs, miRNA* and hairpin sequences from miRbase and miRNAs from RNAdb), piRNA sequences (from RNAdb), Rfam sequences, Fantom3 sequences (mouse non-coding RNAs) and other non-coding RNA databases in the small RNA libraries from the first sequencing run (Cambridge Research Institute run). The percentage of unannotated reads that align to the genome is shown.

Small RNA Library	% miRNA reads total	% piRNA reads	% Rfam reads	% Fantom3 reads	% Hinv, Evofold, Antisense ncRNA pipeline and literature data set reads	% Un-annotated genomic reads
Stage 8	0.15	0.35	1.45	2.47	0.35	95.2
Stage 10	0.27	0.33	1.79	1.27	0.21	96.1
Stage 18	1.21	0.56	1.94	2.25	0.29	93.8
Animal	0.49	1.04	3.13	4.91	0.21	93.2
Vegetal	0.25	0.70	2.94	4.91	0.29	90.9

Table 4.5 Technical repeat of small RNAome annotation (LRI sequencing run)
Annotation of the small RNAome as described in Table 4.4 for the technical repeat sequencing run performed at the Cancer Research UK London Research Institute about one year after the first sequencing run at the Cambridge Research Institute.

4.2.2.1 MicroRNA characterisation

In order to gain an accurate profile of microRNA expression in the five small RNA libraries, trimmed reads of length 22-23 nt with a perfect match to a miRbase microRNA tag which was sequenced a minimum of 10 times (i.e. 10 reads) in at least one of the libraries were retained. A read length of 22-23 nt was selected to eliminate noise from shorter reads aligning to part of a miRNA sequence, or multiple miRNA sequences. 169 miRbase microRNA sequences passed this filter including 51 known *X. tropicalis* and two *X. laevis* miRNAs, xla-miR-427 and xla-miR-449. The filter vastly cuts down the number of microRNA tags from a maximum of 851 microRNA tags shown in Figure 4.3.

The five most frequently sequenced microRNAs in each library are presented in Table 4.6. Small RNA sequencing is semi-quantitative, therefore sequencing frequency correlates with miRNA abundance (Creighton et al., 2009). Notably, xtr-miR-206 and cfa-miR-30e dominate the stage 18 microRNA profile. xtr-miR-206 reads are 33 % and cfa-miR-30e reads are 19 % of all miRNA reads at stage 18. hsa-miR-1825 is the most abundant stage 8, stage 10 animal and stage 10 vegetal miRNA, strikingly accounting for 50 % of miRNA reads in the vegetal pole. hsa-miR-1825 is the 7th most abundant miRNA in the stage 10 whole embryo, accounting for 5.3 % of miRNA reads. Comparing the microRNAs sequenced in the stage 10, stage 10 animal and stage 10 vegetal libraries can estimate the reproducibility of the microRNA profiles. As shown in Figure 4.6, 97 microRNAs out of a total of 154 were sequenced in the stage 10 whole embryo, stage 10 animal and stage 10 vegetal libraries. A further 18 microRNAs were sequenced in either the stage 10 animal or vegetal libraries and the stage 10 library, giving a total overlap of microRNAs sequenced in the stage 10 whole embryo and dissected libraries of 115/154 microRNAs, or 75 %. The five microRNAs found in the stage 10 library, but not in the animal or vegetal libraries, were present at read numbers less than 26 in the stage 10 whole embryo, and were detected at low read numbers at stage 8 and/or 18. This implies that the limit of sequencing sensitivity was approached. MicroRNAs sequenced in animal or vegetal libraries but not the stage 10 whole embryo libraries were generally present in the stage 18 libraries. This could be explained by the

extra time required to dissect the animal and vegetal poles at stage 10, which was performed after freezing the whole embryo samples.

Heat maps of microRNA expression in the five small RNA libraries were generated by multiple sequence alignment of all microRNA tags that passed a read length filter of 22-23 nt and zero mismatches in sequence to a miRbase microRNA to create a dendrogram. The dendrogram divides into four branches based on microRNA sequence similarity (Figure 4.7). MicroRNA expression was illustrated using a colour scale of log (base 10) of normalised read number in each library, where green corresponds to low normalised read numbers and red to high normalised read numbers. The most abundant, dynamically expressed and spatially localised microRNAs in the dataset can be selected from the heat maps.

In Branch 1 (Figure 4.8), xtr-miR-101a is expressed at comparatively high levels in all the libraries (~200 reads at stage 8 and stage 10, and over 800 reads at stage 18), whereas xtr-miR-130b is the most abundant and dynamic miRNA, increasing almost 200-fold in abundance between stage 10 and stage 18. xtr-miR-183, xtr-miR-130c and xtr-miR-130b are the most dynamically expressed miRNAs, increasing in expression between stage 10 and stage 18. dme-miR-312 is the most localised miRNA, showing 25-fold higher abundance in the animal pole than the vegetal pole at stage 10.

In Branch 2 (Figure 4.9), cfa-miR-363 and hsa-miR-363 are the most abundant microRNAs, and are dynamic in expression, increasing over 14-fold and 45-fold respectively from stage 10 to stage 18. xtr-miR-367 is highly abundant at stage 18, and is the most dynamically expressed microRNA increasing from zero reads at stage 10 to 124 reads at stage 18. xtr-miR-107 is the most spatially localised miRNA with ~8-fold higher abundance in the animal pole compared to the vegetal pole at stage 10.

dme-miR-184 is the most abundant miRNA in branch 3 of the heatmap (Figure 4.10), and is the most spatially localised with ~13-fold higher abundance in the animal pole than the vegetal pole at stage 10. In contrast, dme-miR-263a and xtr-miR-24a are more abundant in the vegetal pole. gga-miR-146c is the most dynamically expressed miRNA in Branch 3, again increasing in abundance between stage 10 and stage 18.

Branch 4 of the heat map (Figure 4.11) reveals that hsa-miR-1825 is the most stably expressed abundant miRNA, whereas the most abundant miRNA sequenced in the dataset, xtr-miR-206 shows dynamic expression by increasing over 300-fold between stages 10 and 18. cfa-miR-30e is a highly abundant miRNA that increases 12-fold between stage 10 and stage 18. The let-7 family of miRNAs is also highly abundant at stages 10 and 18. xtr-miR-16c, mmu-miR-466i, rno-miR-10b-3p, mmu-miR-10b, xtr-miR-10b and cfa-miR-200b follow a trend of increasing abundance at stage 18 compared to stages 8 and 10. Spatially, bmo-miR-184 is the most localised miRNA, with 10-fold higher normalised reads in the animal compared to the vegetal pole at stage 10. The spatial and dynamic expression profiles of miRNAs will be experimentally validated in section 4.2.4.2.

As previously stated, small RNA libraries were re-sequenced to obtain a technical repeat. For this dataset, the same filter was applied to microRNA tags that matched a miRbase miRNA. On re-sequencing, 179 microRNAs passed the filter, compared to 169 with the first sequencing run. 23 new microRNAs were present in the re-sequenced dataset (Table 4.7), however, 13 microRNAs present in the original dataset were not re-sequenced meaning that the technical repeat was 87 % identical to the original dataset in terms of microRNA tags. The new microRNAs sequenced in the technical repeat were all relatively low read numbers; the most abundant microRNA, xtr-miR-148b was sequenced 29 times at stage 18.

Since miRbase has been updated to include new microRNAs since this work began, the microRNA sequences in miRbase 17 were filtered to remove identical entries from different species and compared to the trimmed, aligned, small RNA sequencing reads. miRbase 17 contains 12,187 miRNA tags compared to the 5,213 tags in miRbase 14 after filtering. When compared to miRbase 17, a total of 414 additional microRNAs were found in the small RNA libraries that were not listed in miRbase 14, bringing the total number of *Xenopus* microRNAs I identified to 606. New miRbase 17 microRNAs found in the libraries are listed in the Appendix (Table 6.1).

Small RNA library	MicroRNA	Normalised Read Number	% of total miRNA reads
Stage 8	hsa-miR-1825	991	18.2
	hsa-miR-423-5p	423	7.8
	xtr-let-7c	267	4.9
	xtr-let-7f	255	4.7
	xtr-miR-148a	234	4.3
Stage 10	hsa-miR-423-5p	1482	10.1
	cfa-miR-30e	1042	7.1
	dme-miR-184	1018	7.0
	bmo-miR-184	996	6.8
	xtr-let-7f	785	5.4
Stage 18	xtr-miR-206	21374	33.4
	cfa-miR-30e	12401	19.4
	mmu-miR-10b	2857	4.5
	rno-miR-10a-3p	2856	4.5
	xtr-miR-130b	2387	3.7
Stage 10 animal	hsa-miR-1825	2624	12.0
	dme-miR-184	2374	10.9
	bmo-miR-184	2351	10.8
	cfa-miR-30e	1379	6.3
	hsa-miR-423-5p	1347	6.2
Stage 10 vegetal	hsa-miR-1825	4952	49.7
	cfa-miR-30e	818	8.2
	hsa-miR-423-5p	417	4.2
	xtr-let-7f	257	2.6
	bmo-miR-184	231	2.3

Table 4.6 The most frequently sequenced *X. tropicalis* miRNAs in early development

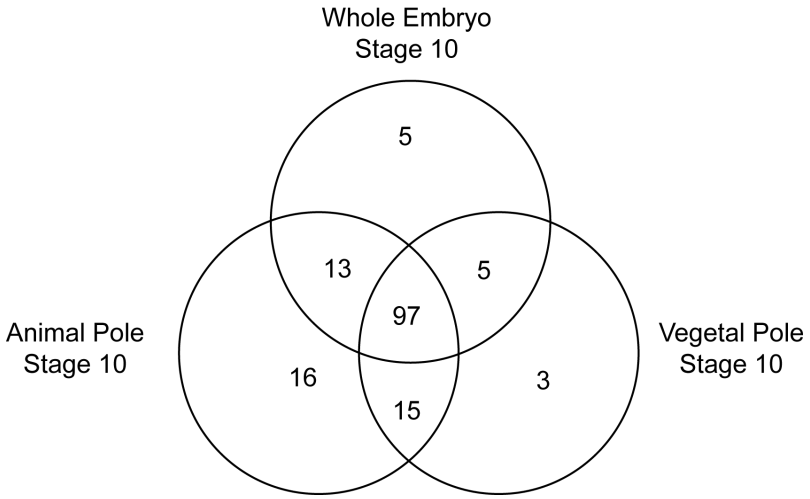


Figure 4.6 Venn diagram showing overlap of miRNA sequencing at stage 10

MicroRNAs that were sequenced in any of the stage 10 small RNA libraries (whole embryo, animal or vegetal pole) are plotted to show the number of microRNAs that are present in more than one stage 10 library.

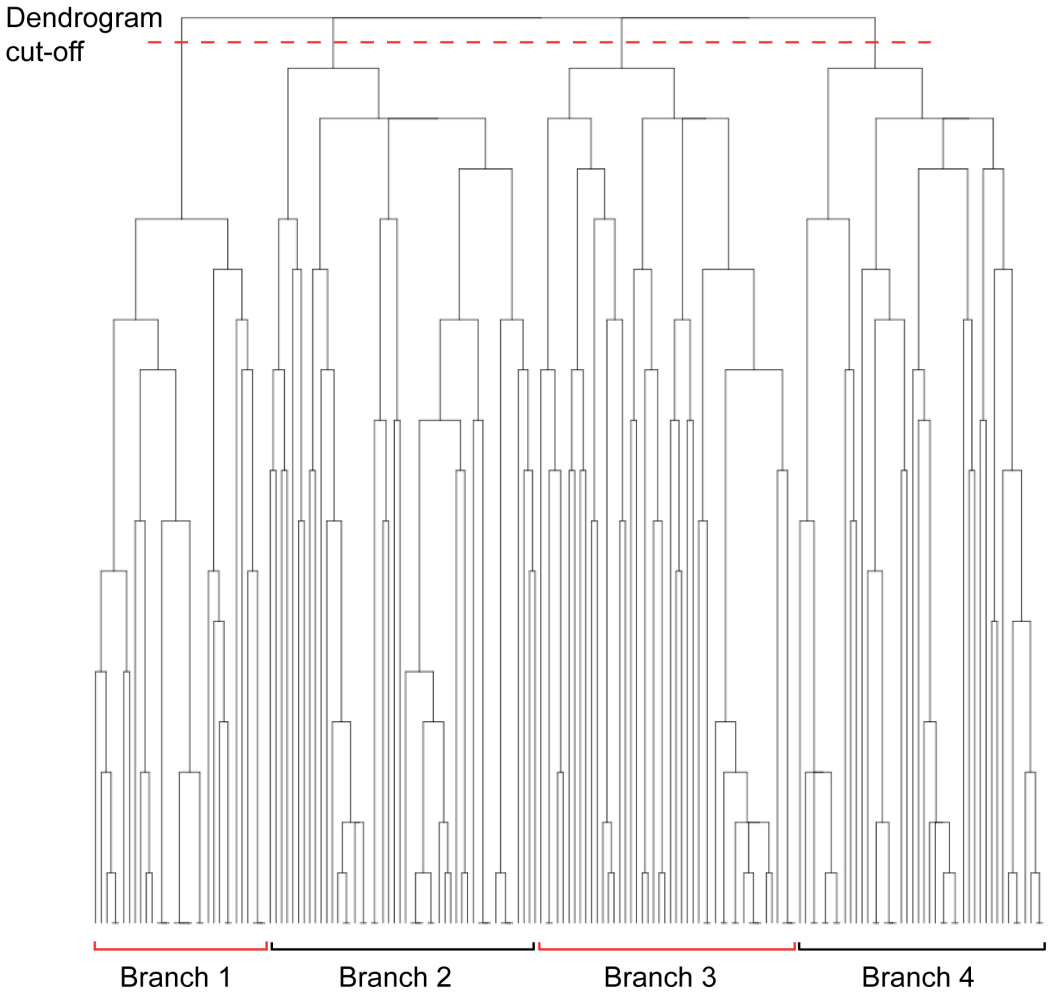


Figure 4.7 Dendrogram of miRNAs for heat map expression profiles

MicroRNA tags in all five libraries passing a filter of 22-23 nt read length and zero mismatches to a miRbase microRNA were aligned and grouped by sequence similarity to form a dendrogram with four major branches. There are 169 microRNAs in total.

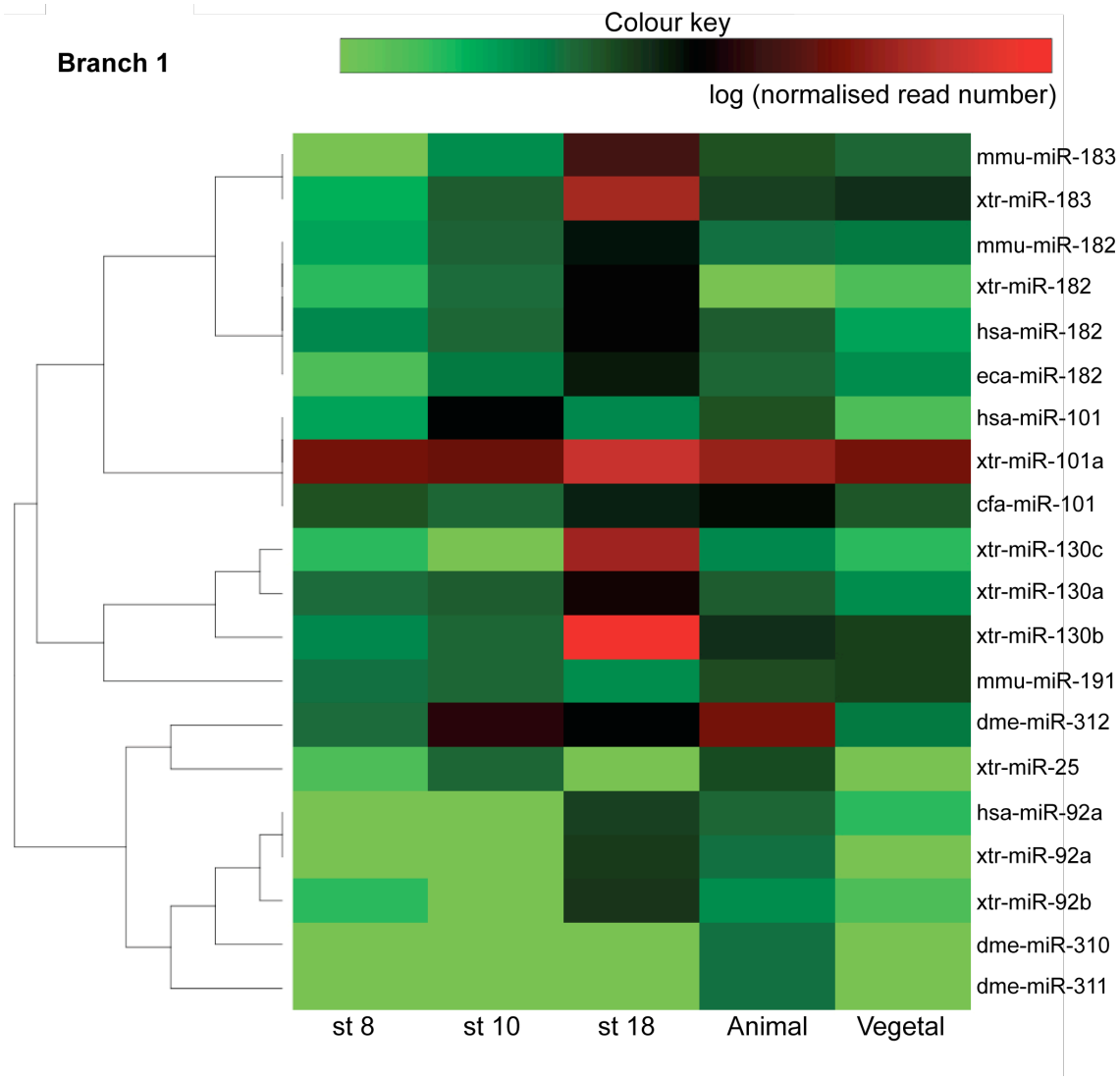


Figure 4.8 MicroRNA heat map for dendrogram branch 1 showing dynamic and localised miRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high miRNA read numbers respectively. MicroRNA tags of 22-23 nt length with a perfect match to a miRbase microRNA sequence were analysed.

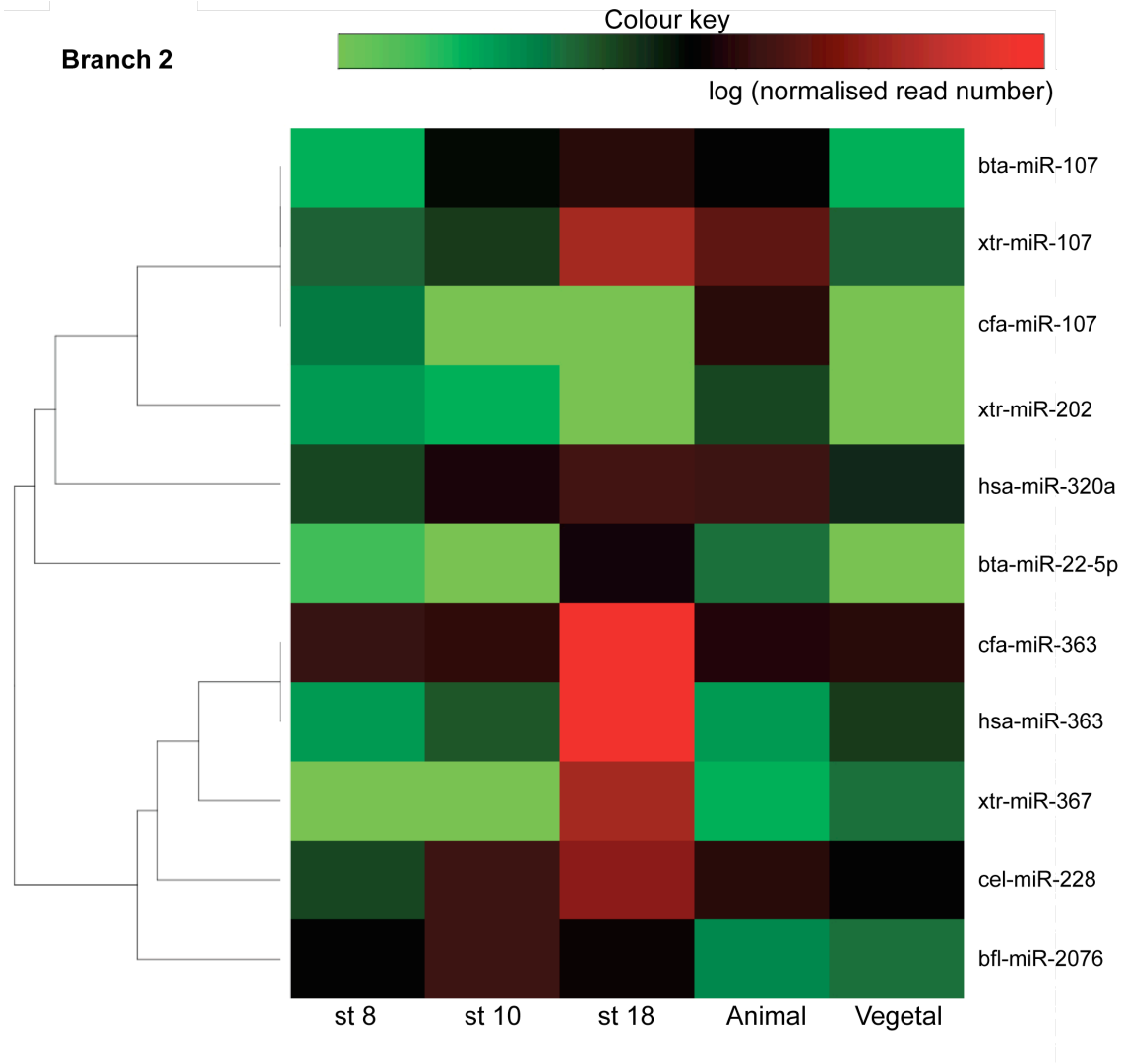


Figure 4.9 MicroRNA heat map for dendrogram branch 2 showing dynamic and localised miRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high miRNA read numbers respectively.

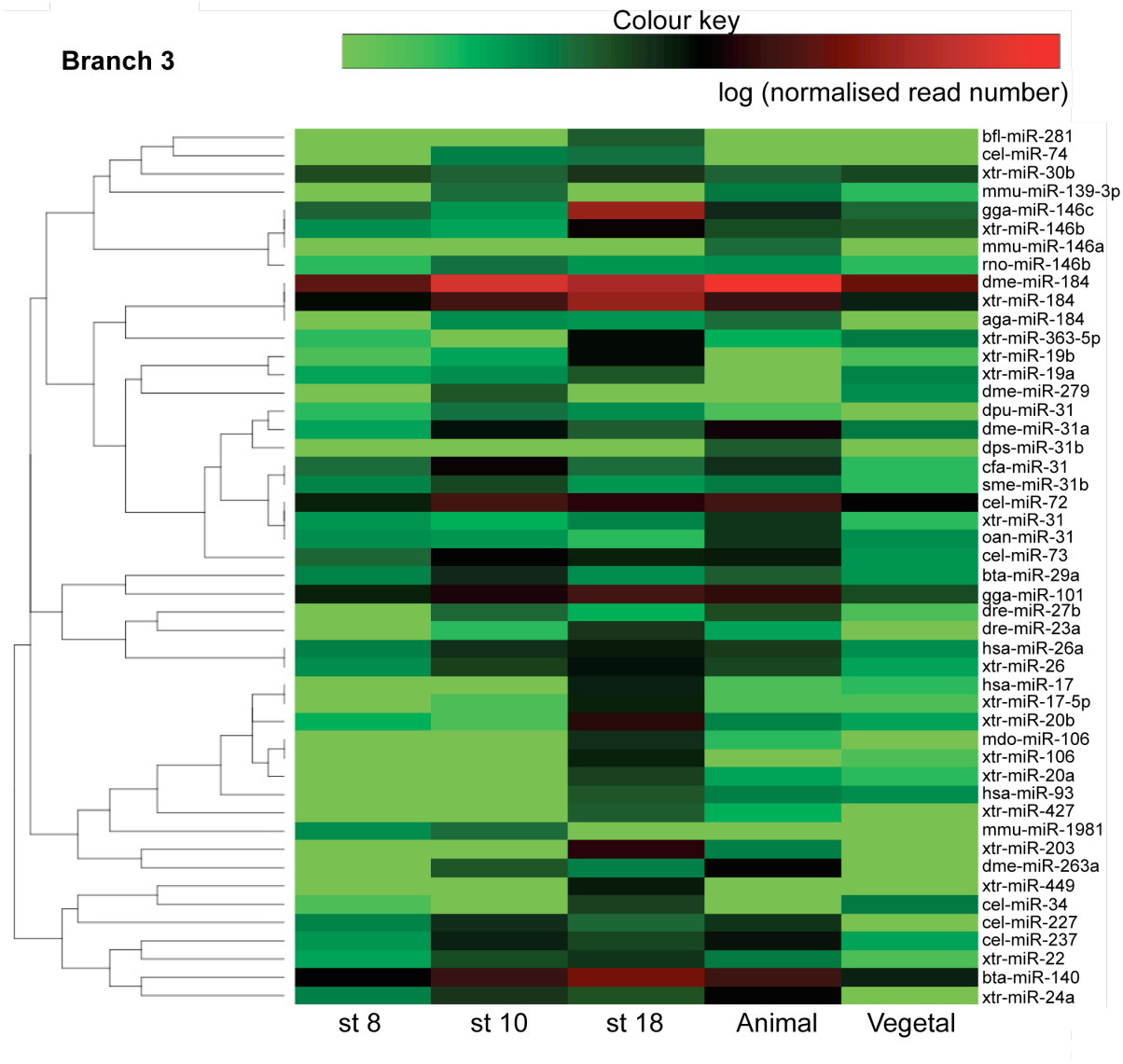


Figure 4.10 MicroRNA heat map for dendrogram branch 3 showing dynamic and localised miRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high miRNA read numbers respectively.

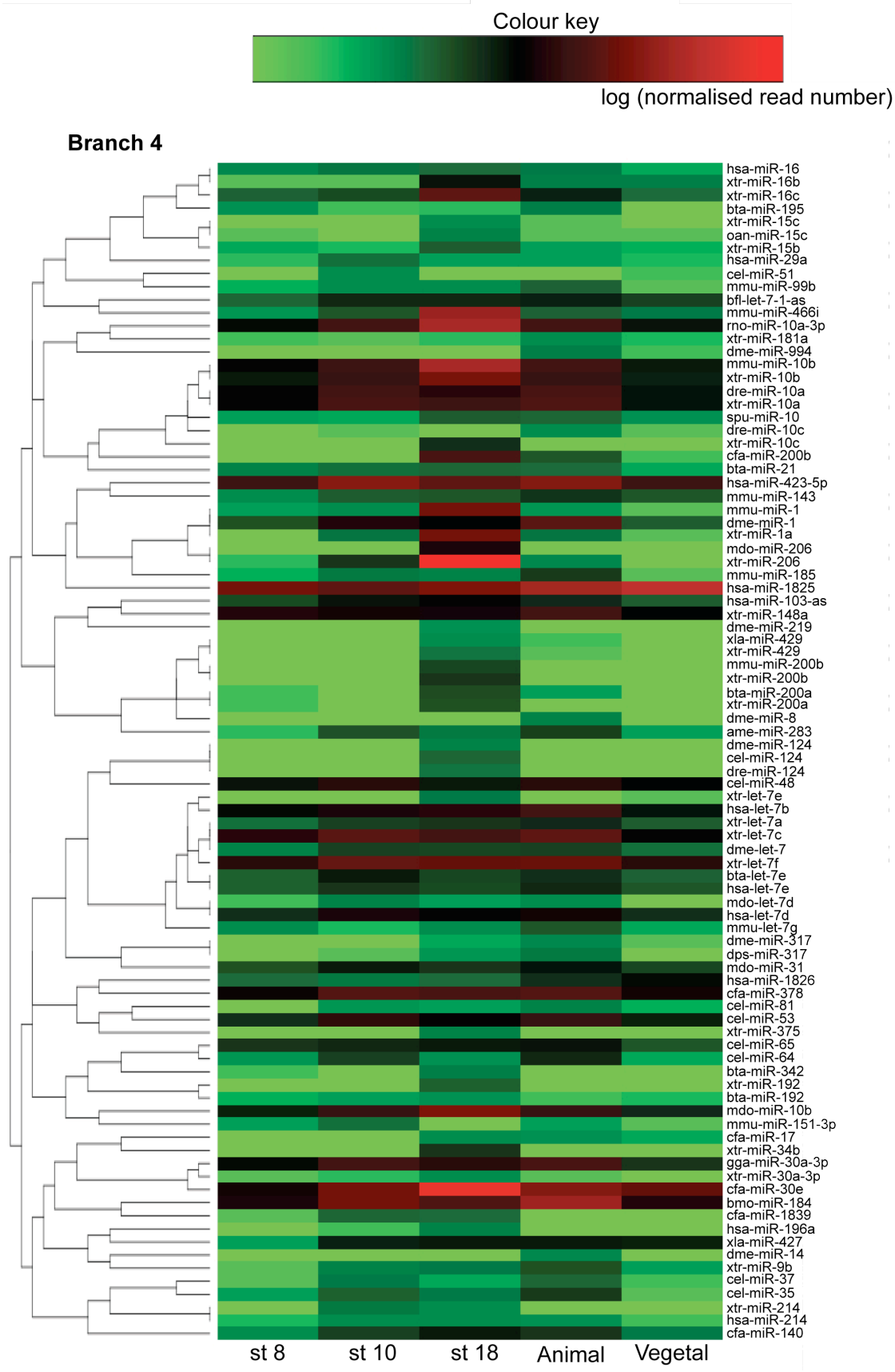


Figure 4.11 MicroRNA heat map for dendrogram branch 4

MicroRNA	Stage 8	Stage 10	Stage 18	Stage 10 animal	Stage 10 vegetal
xtr-miR-148b	17	0	29	16	7
dme-miR-306	0	0	19	0	0
mmu-miR-29c	0	13	18	9	6
mml-miR-96	0	0	16	0	0
hsa-miR-224	0	0	15	0	0
dre-miR-456	0	17	13	0	9
cfa-miR-181a	0	1	13	0	0
dmo-miR-317	0	0	13	22	2
xtr-miR-34a	0	0	13	9	0
xtr-miR-7	2	1	12	0	2
rno-miR-143	4	10	11	1	5
xtr-miR-15a	8	13	9	6	2
oan-miR-30d	2	0	8	19	8
bta-miR-30a-5p	2	13	5	7	3
mmu-miR-744	13	0	5	0	10
oan-miR-30a	6	13	4	13	1
hsa-miR-21	16	20	0	0	1
rno-miR-350	0	18	0	0	0
hsa-miR-26b	0	17	0	0	0
mmu-miR-15b	0	12	0	0	5
dme-miR-9c	0	0	0	14	1
hsa-miR-28-3p	0	0	0	14	0
hsa-miR-532-5p	0	0	0	20	0
mmu-let-7i	0	0	0	14	0

Table 4.7 New microRNAs obtained by re-sequencing the libraries

Normalised read numbers for miRNAs that were obtained by re-sequencing the small RNA libraries are shown (LRI technical repeat sequencing run).

4.2.2.2 Piwi-interacting RNA characterisation

To obtain profiles of piRNA expression and localisation in the small RNA libraries, trimmed reads of length 25-30 nt that perfectly matched a piRNA tag from RNAdb with at least 10 reads in one small RNA library were retained. 64 piRNA sequences passed this filter and are listed in Figure 4.13-4.16. Table 4.8 shows the five most frequently sequenced piRNAs in each small RNA library and the percentage of piRNA reads they represent. Strikingly, PIR32037 is the most abundant piRNA in all five small RNA libraries, and dominates the piRNA expression profile of the whole embryo libraries, comprising 71.6 % to 87.8 % of all the piRNA reads. Interestingly, the PIR32037 read percentage is lower (44.3 %) in the stage 10 vegetal pole library because read numbers of other piRNAs such as PIR75449, PIR72436, PIR75451 and PIR63462 are higher in the vegetal pole. This suggests that the vegetal pole has a more diverse piRNA population than the animal pole and whole embryo at stages 8 to 18.

A dendrogram shows the sequence similarity when all piRNA tags passing filtering criteria are aligned and the four major branches shown (Figure 4.12) are analysed by heatmap expression profiling. In Branch 1 of the heat map (Figure 4.13), PIR32037 stands out as the most abundant piRNA and is stably expressed from stages 8 to 18 and is not localised to the animal or vegetal pole at stage 10. Most piRNAs are relatively low read number (green on the colour scale) with the exception of PIR63462, PIR63461 and PIR63463 which are all present at over 300 reads in more than one small RNA library.

In Branch 2 of the piRNA heatmap (Figure 4.14), PIR31623 and PIR64428 are highly abundant and stably expressed, whereas PIR57010 increases 18-fold between stage 10 and stage 18. PIR 5416 is 9.7-fold enriched in the vegetal pole compared to the animal pole at stage 10.

PIR63554 is the most abundant piRNA in Branch 3 of the heatmap (Figure 4.15) and is vegetally enriched 7.7-fold compared to the animal pole at stage 10. PIR59215 appears to be dynamically expressed with expression peaking at stage 10, whereas most piRNAs on this branch are stably expressed at read numbers of 10 or less. Interestingly,

PIR32354 is enriched 8.4-fold in the animal pole and PIR37983 is enriched 8-fold in the vegetal pole at stage 10.

PIR75451, PIR75449 and PIR72436 are the most abundant piRNAs in Branch 4 (Figure 4.16) and show strong vegetal pole enrichment with over 1000 reads in the vegetal pole library. PIR75451 and PIR64874 expression peaks at stage 10, whereas PIR75449, PIR75450, and PIR32103 peak at stage 18. Experimental validation of dynamic and spatial piRNA expression is detailed in section 4.2.4.3.

Small RNA library	piRNA	Normalised Read Number	% of total piRNA reads
Stage 8	PIR32037	9101	87.8
	PIR63461	323	3.1
	PIR63463	139	1.3
	PIR63462	105	1.0
	PIR64428	104	1.0
Stage 10	PIR32037	8396	83.4
	PIR63461	348	3.5
	PIR75451	159	1.6
	PIR63463	117	1.2
	PIR64874	112	1.1
Stage 18	PIR32037	7102	71.6
	PIR63462	359	3.6
	PIR32103	236	2.4
	PIR63463	235	2.4
	PIR75200	139	1.4
Stage 10 animal	PIR32037	9579	74.8
	PIR63462	752	5.9
	PIR64428	520	4.1
	PIR63463	192	1.5
	PIR72436	163	1.3
Stage 10 vegetal	PIR32037	7198	44.3
	PIR75449	1536	9.5
	PIR72436	1362	8.4
	PIR75451	1263	7.8
	PIR63462	866	5.3

Table 4.8 The most frequently sequenced piRNAs in early development

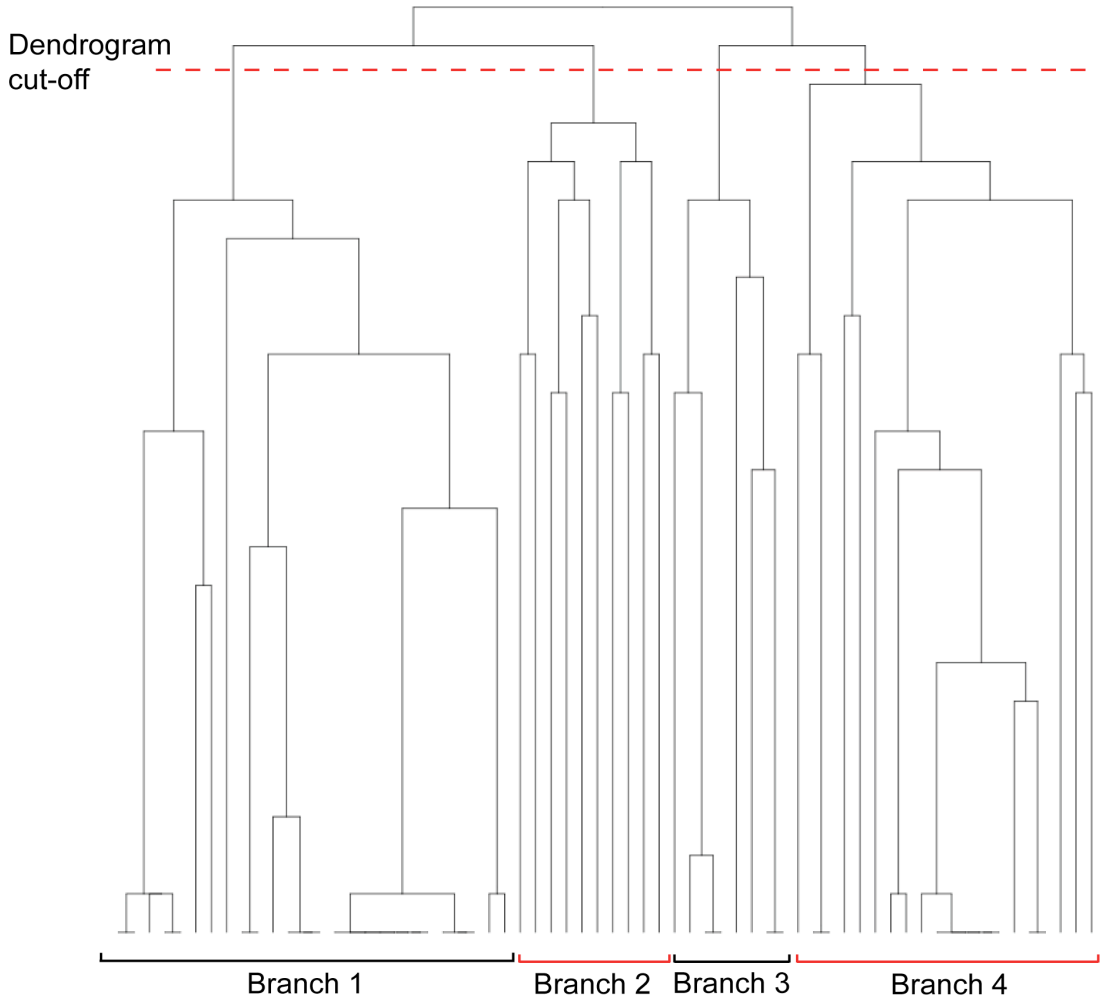


Figure 4.12 Dendrogram of piRNAs for heat map expression profiling

piRNA tags in all five libraries passing a filter of 25-30 nt read length and zero mismatches to an RNAdb piRNA were aligned and grouped by sequence similarity to form a dendrogram with four major branches. There are 64 piRNAs in total.

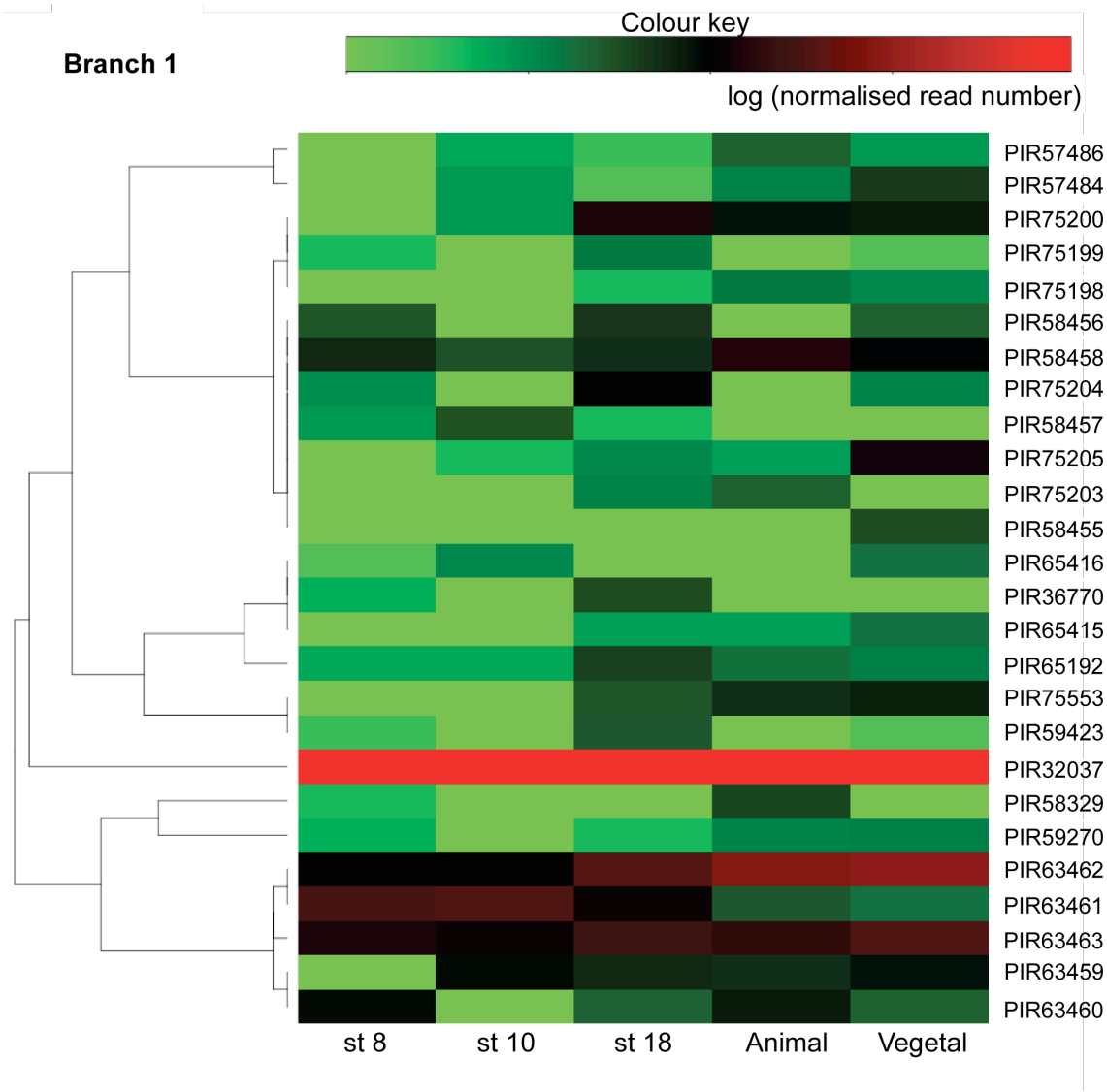


Figure 4.13 piRNA heat map for dendrogram branch 1 showing dynamics and localisation of piRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high piRNA read numbers respectively. piRNA tags of length 25-30 nt match a piRNA sequence in RNAdb with no mismatches.

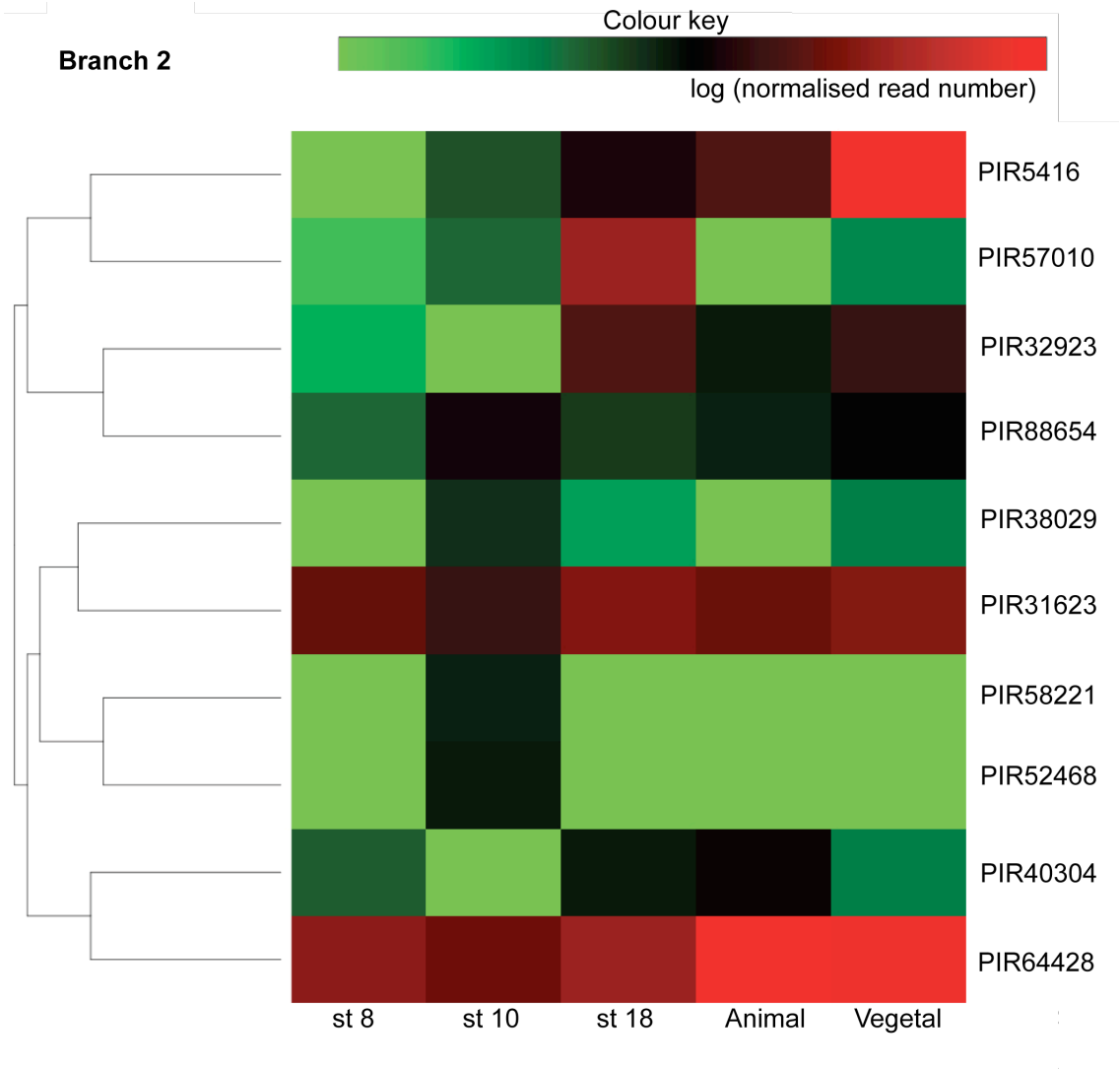


Figure 4.14 piRNA heat map for dendrogram branch 2 showing dynamics and localisation of piRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high piRNA read numbers respectively.

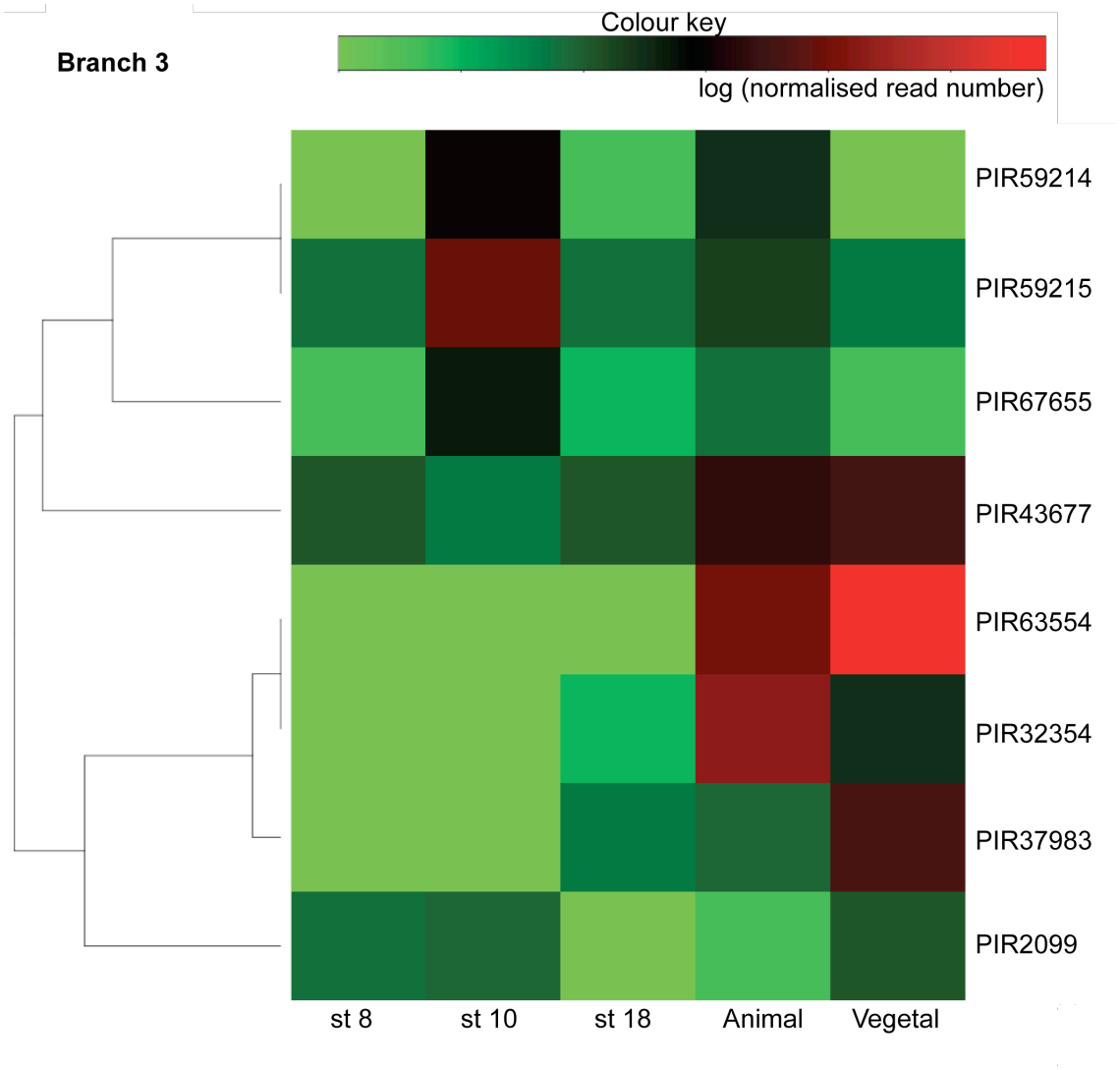


Figure 4.15 piRNA heat map for dendrogram branch 3 showing dynamics and localisation of piRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high piRNA read numbers respectively.

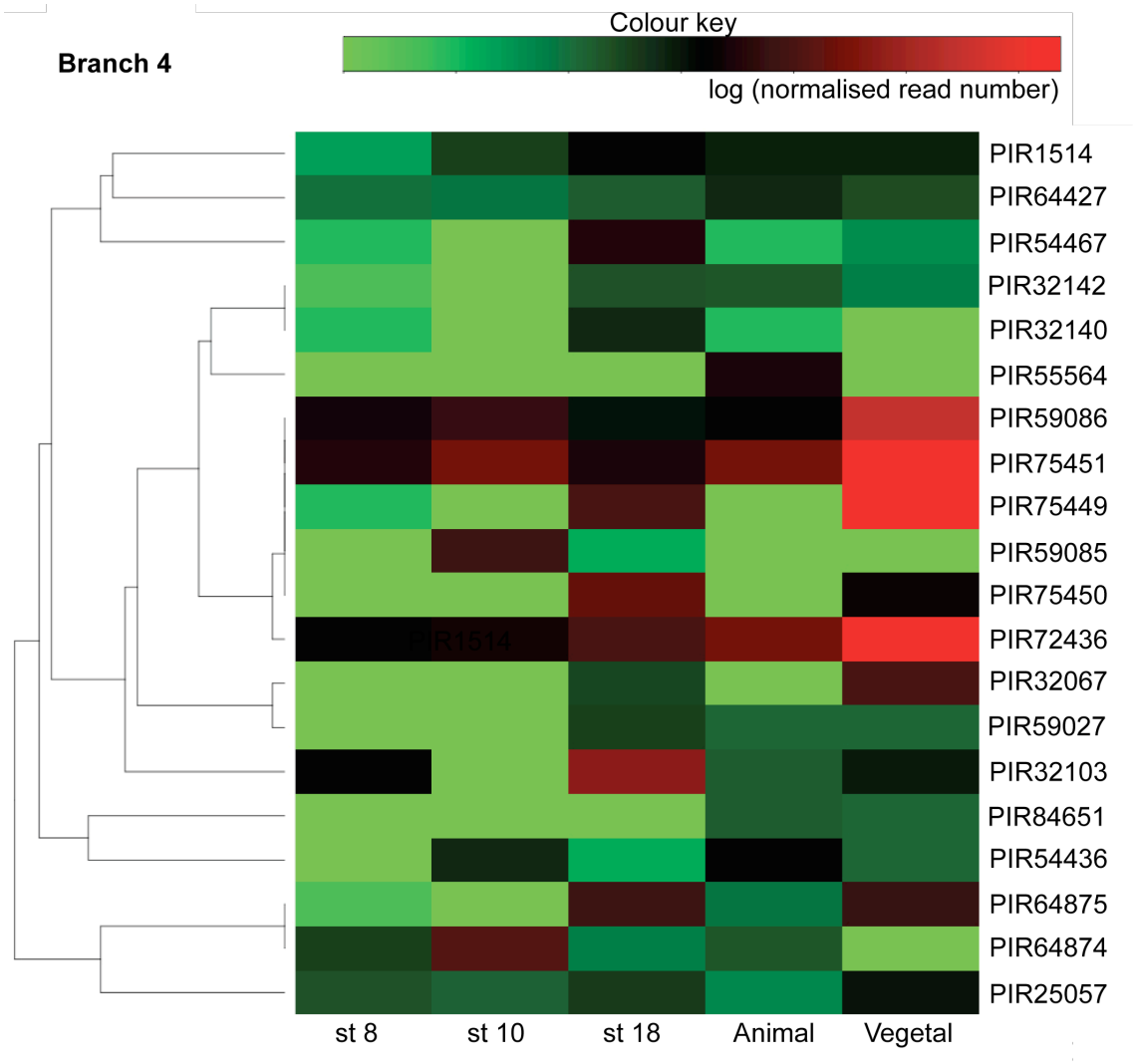


Figure 4.16 piRNA heat map for dendrogram branch 4 showing dynamics and localisation of piRNA expression

Heat maps were calculated using a colour key based on log base 10 of values of normalised read numbers. Green corresponds to low and red to high piRNA read numbers respectively.

4.2.3 MicroRNA and Piwi-interacting RNA expression show contrasting dynamics and localisation in the early embryo

MicroRNA and piRNA tags that passed the read length and perfect match criteria were next analysed for their expression profiles from blastula stage (stage 8) to late neurula stage (stage 18) and in the animal and vegetal pole at the onset of gastrulation (stage 10) to investigate the dynamics and localisation of microRNA and piRNA expression in early *X. tropicalis* development.

MicroRNA and piRNA read numbers at stage 8, stage 10 and stage 18 are shown relative to stage 8 in Figure 4.17. MicroRNAs increase in abundance 3.7-fold on average between stage 8 and stage 10 and 75-fold between stage 8 and stage 18. MicroRNA expression is therefore characterised by a dramatic increase at stage 18, although individual microRNAs in the dataset do decrease in abundance at stages 10 and 18 (Figure 4.17). There are 13 microRNAs which increase over 100-fold between stage 8 and stage 18, including the most frequently sequenced stage 18 microRNAs, xtr-miR-206 and cfa-miR-30e (Table 4.6). This agrees with the previous conclusion that a small number of highly abundant miRNAs dominate the stage 18 microRNA profile. In contrast, piRNA expression is less dynamic in early development, increasing 6.8-fold at stage 10 relative to stage 8 on average, and 15-fold at stage 18 relative to stage 10.

MicroRNA and piRNA normalised read numbers in the animal and vegetal pole are shown in Figure 4.18. MicroRNAs are distributed towards the animal pole (Figure 4.18A), indicating that in general, microRNAs are more abundant in the animal pole compared to the vegetal pole. piRNA localisation is more variable in the embryo, with no bias towards expression in either the animal or vegetal pole. This result is not skewed by the higher cell number in the animal compared to the vegetal pole, as the libraries are normalised to the total number of reads aligning to the genome (i.e. total number of small RNAs cloned) in each library, and contrasting results were obtained for microRNA and piRNA localisation.

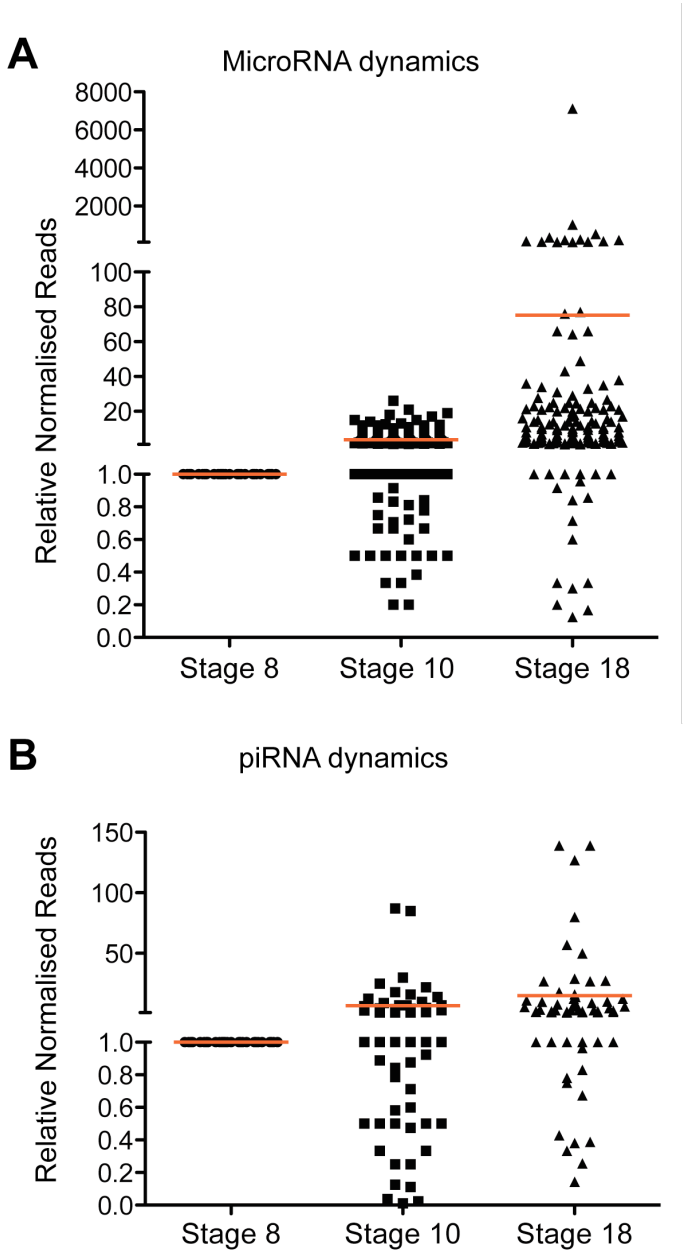


Figure 4.17 MicroRNA and piRNA dynamics from the blastula to late neurula stages of *X. tropicalis* development

A. Normalised read numbers for individual microRNAs at stages 10 and 18 are shown relative to stage 8 values for the stage 8, stage 10 and stage 18 libraries. Essentially, all stage 8 reads are set to a value of 1. The mean is indicated by a red line. Note that the axes are continuous, i.e., values greater than 1 and less than or equal to 100 are plotted on the middle section of the graph and the upper section contains values greater than 100 and less than or equal to 8000. **B.** Piwi-interacting RNA read numbers are displayed as above.

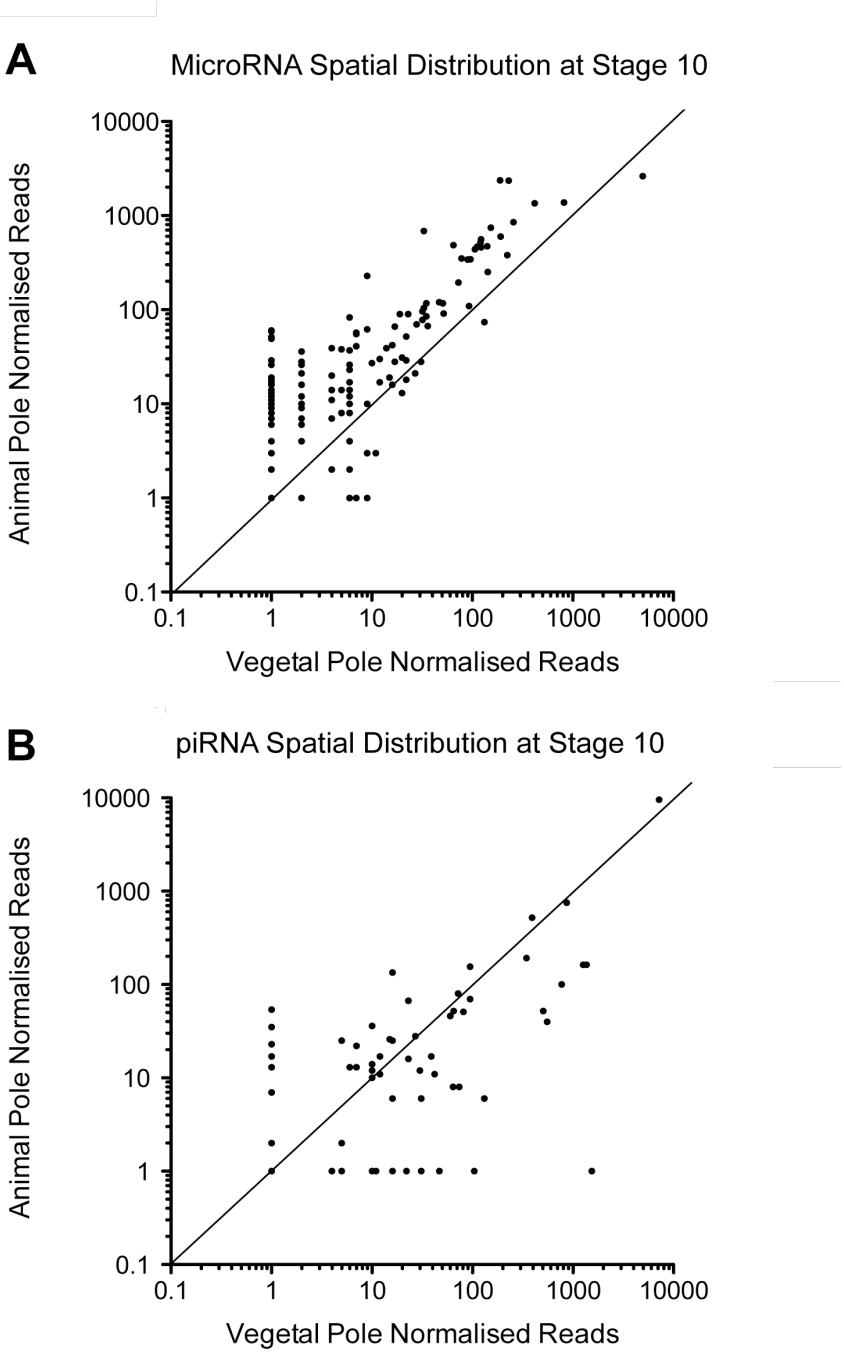


Figure 4.18 Spatial distribution of microRNAs and piRNAs at gastrulation shows that microRNAs are more localised to the animal pole at stage 10.

A. Normalised microRNA reads in the stage 10 animal and stage 10 vegetal libraries are plotted on a logarithmic scale with y = x superimposed to show the animal-vegetal distribution. **B.** Normalised piRNA reads are plotted as above.

4.2.4 Experimental validation of small RNA-seq data

4.2.4.1 Optimisation of small RNA detection

In order to experimentally validate the expression profiles of small RNAs in the libraries, sensitive detection techniques were required. I optimised microRNA qPCR and small RNA Northern blot protocols. For detection of miRNAs by qPCR as shown in Figure 4.19A, total RNA is polyadenylated and reverse transcription is performed using the NVT-adapter primer that hybridises to the 5' poly-adenosines (Shi and Chiang, 2005). The resulting cDNA is used as the template for miRNA qPCR amplification, with a forward primer identical to the mature miRNA sequence and a reverse primer in the adapter sequence. The predicted sizes of PCR products according to the miRNA-qPCR protocol for two *X. tropicalis* small RNAs, xtr-miR-427 and U6 were 69 bp and 98 bp respectively. This was confirmed by native gel electrophoresis of the miRNA RT-PCR products (Figure 4.19B).

MicroRNA qPCR profiles for two known *Xenopus laevis* microRNAs that are present in early development, xla-miR-427 and xla-miR-16c (Watanabe et al., 2005) were consistent with those previously described (Figure 4.20A), and the qPCR profile of xla-miR-427 mirrored the expression shown by Northern blot in Figure 4.20B. Relative miRNA levels to *ornithine decarboxylase (ODC)* or the spliceosomal small RNA *U6* were calculated. *ODC* was selected because its expression levels are constant throughout *Xenopus* development (Paris et al., 1988). U6 was also chosen because it is uniformly expressed during *Xenopus* development (data not shown) and is an internal control for polyadenylation efficiency of small RNAs. There was no difference in results when normalisation to *ODC* or *U6* was compared.

Northern blotting with UV cross-linking such as that shown in Figure 4.20B requires large quantities of total RNA for effective detection of small RNAs (30 μg), which is impractical for large-scale validation of multiple small RNA species in *X. tropicalis.* An alternative chemical cross-linking method using 1-ethyl-3-3-dimethylaminopropyl carbodiimide, (EDC), a chemical that cross-links the 5' phosphate of small RNAs to the nylon membrane (Pall and Hamilton, 2008) was compared to UV cross-linking. As shown in Figure 4.21, EDC cross-linking strongly increases the mature xla-miR-427

signal so that a signal can be detected with as little as 5 μg of RNA, but has no significant effect on the pre-miRNA signal.

4.2.4.2 MicroRNA validation

I used microRNA qPCR to confirm the existence of the sequenced microRNAs and validate their dynamic and spatial expression patterns. Only microRNAs with qPCR dissociation curves indicating a single PCR product were retained for analysis and miRNA forward primers were screened for cross-reactivity with other similar microRNAs. Experimental validation of dynamically expressed microRNAs selected from Figure 4.8-4.11 is presented in Figure 4.23 and 4.33. The small RNA seq read number profiles closely matched the qPCR profile at stages 8, 10 and 18 for the following miRNAs: cfa-miR-30e, xtr-miR-16c, xla-miR-427, xtr-miR-101a, xtr-miR-182, xtr-miR-206 and xtr-miR-148a (Figure 4.22). The qPCR profiles for xtr-let-7c, xtr-let-7f and xtr-miR-10a showed lower miRNA abundance at stage 10 than expected from the small RNA seq read profile (Figure 4.23). This is not due to a quality issue with the stage 10 small RNA library or qPCR input RNA, as strong miRNA expression at stage 10 is validated by qPCR for xla-miR-427. The stage 8 miRNA level for xtr-miR-184 is lower than expected in the qPCR compared to the small RNA sequencing. In short, 7 out of 11 microRNAs experimentally validate at all stages and there is a discrepancy between the small RNA seq and experimental validation at one stage for 4 out of 11 miRNAs assessed. The validation rate for dynamic microRNAs is therefore 64 %.

Next, spatially localised microRNAs were experimentally validated (Figure 4.24). xla-miR-427 was selected as a control microRNA that is abundant at similar levels in the animal and vegetal pole by small RNA sequencing. The microRNA qPCR for xla-miR-427 was similar to the small RNA sequencing read profile. MicroRNAs with qPCR profiles where the animal and vegetal abundances validated the small RNA sequencing results included bta-miR-140, gga-miR-30a-5p and hsa-mir-1825. Animal and vegetal miRNA levels were within the error bars for miRNA qPCR on hsa-let7b and hsa-miR-423-5p targets, which could indicate equal or weakly localised miRNA levels in the animal or vegetal pole at stage 10. Opposite animal and vegetal miRNA levels compared to the small RNA sequencing were found for dme-miR-184 and xtr-miR-148a by qPCR. In summary, 4/8 spatial miRNAs were successfully validated, a

spatial validation rate of 50 %. The experimental validation of the microRNA expression profiles could be less than 100 % due to bias in the PCR amplification of the small RNA cDNA libraries (see Figure 4.1), which reproducibly over-amplifies certain small RNA sequences (Linsen et al., 2009).

I then investigated the spatial localisation of the most abundant microRNA sequenced, xtr-miR-206. Homologues of miR-206 in other vertebrate species play a role in myogenesis; for example miR-206 regulates skeletal muscle development in the mouse (Anderson et al., 2006), and is induced during muscle development in humans (Rao et al., 2006). Zebrafish miR-206 is expressed in muscle by WISH (Wienholds et al., 2005), therefore a muscle-specific expression pattern was also expected for xtr-miR-206 in *Xenopus*. MicroRNAs can be detected by WISH with DIG end-labelled locked nucleic acid (LNA) probes (Kloosterman et al., 2006). Successful end-labelling of probes was confirmed by blotting for DIG as described in section 2.4.3.1. Figure 4.25 shows that xtr-miR-206 is expressed in presomitic mesoderm at stage 18 in *Xenopus tropicalis* and in somites at stage 24 in *Xenopus laevis*. No specific expression pattern above background staining was observed by WISH for cfa-miR-30e, xtr-miR-130a, xtr-miR-182 and xtr-miR-10b (data not shown). This could be because the protocol was not sufficiently optimised for sensitive microRNA detection, as only the most abundant microRNA (miR-206) was detected.

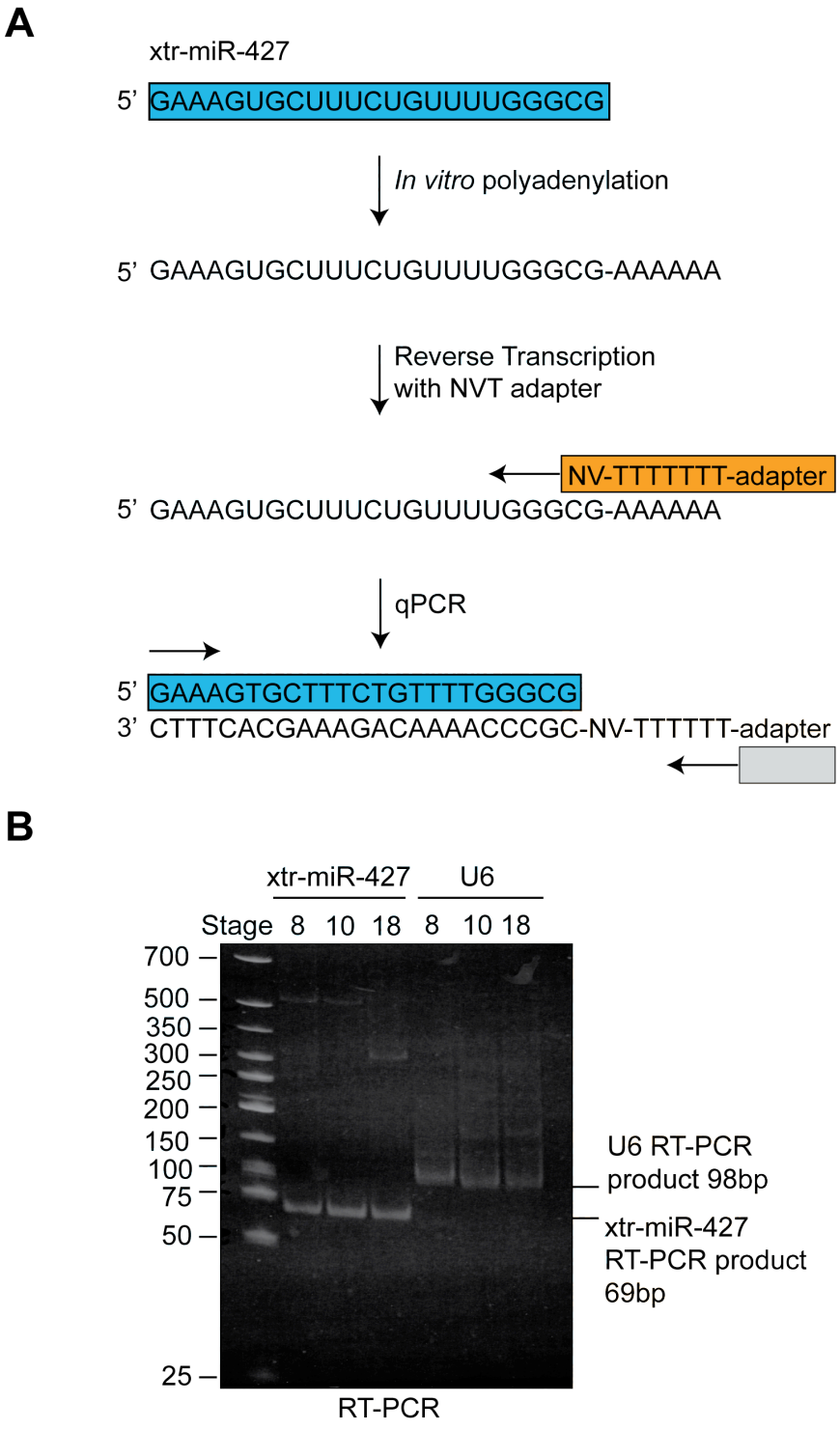


Figure 4.19 microRNA qPCR overview
A. Schematic of miRNA detection by qPCR. Total RNA is polyadenylated and reverse transcribed with the NVT-adapter primer. qPCR is performed with the cDNA template, a forward primer identical in sequence to the microRNA (except Ts substituted for Us) and a reverse primer complementary the adapter sequence. **B.** Native gel electrophoresis of RT-PCR products for xtr-miR-427 and U6 at stages 8, 10 and 18 in *X. tropicalis* (32 cycles). A 6 % native acrylamide gel was run and stained with ethidium bromide.

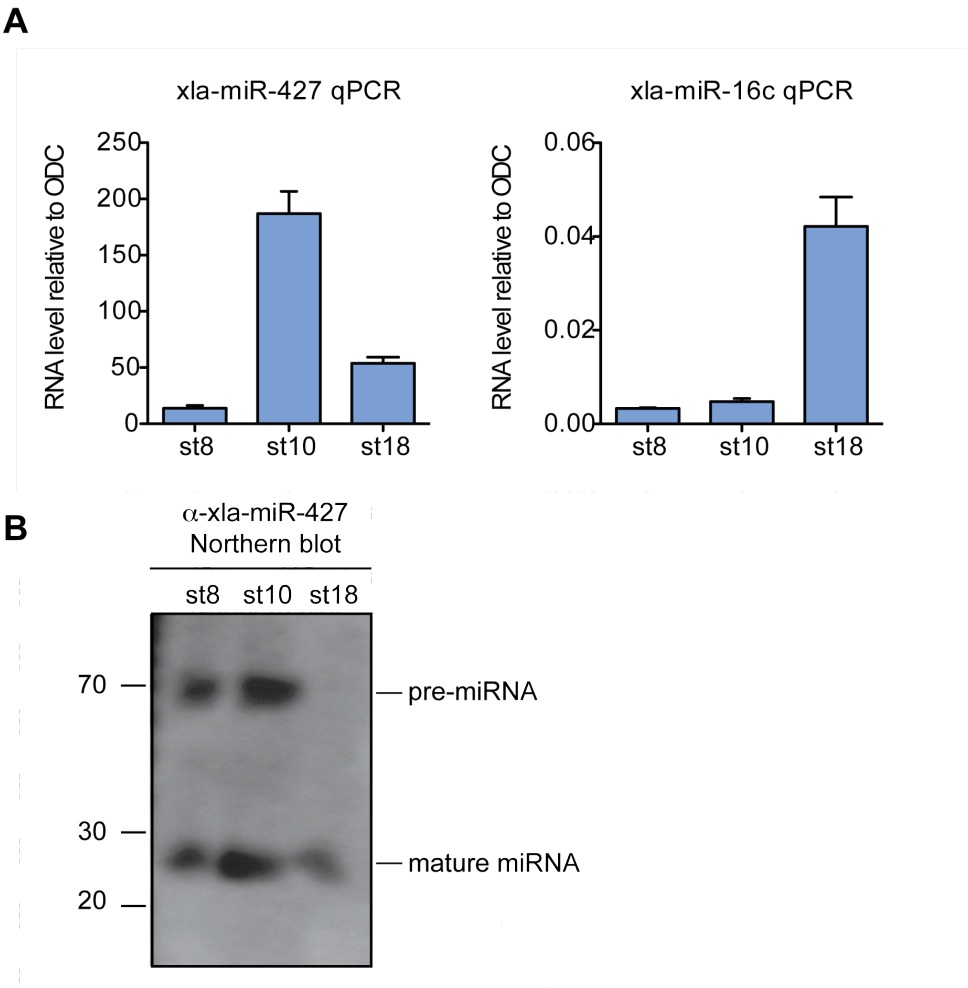


Figure 4.20 MicroRNA qPCR detection of the known *Xenopus laevis* microRNAs xla-miR-427 and xla-miR-16c

A. Total RNA was extracted from whole *X. laevis* embryos at stages 8, 10 and 18 and prepared for miRNA qPCR as described in Figure 4.19. *X. laevis* miRNAs xla-miR-427 and xla-miR-16c were detected by qPCR and relative RNA levels to *Ornithine decarboxylase* (ODC) were calculated. Error bars are standard deviations of triplicate repeats. **B.** Northern blot for xla-miR-427. 30 μg of *X. laevis* total RNA from whole embryos at stages 8, 10 and 18 was probed for xla-miR-427. An antisense probe to the mature miRNA also hybridises to the pre-miRNA. Sizes of molecular weight markers (nt) are indicated.

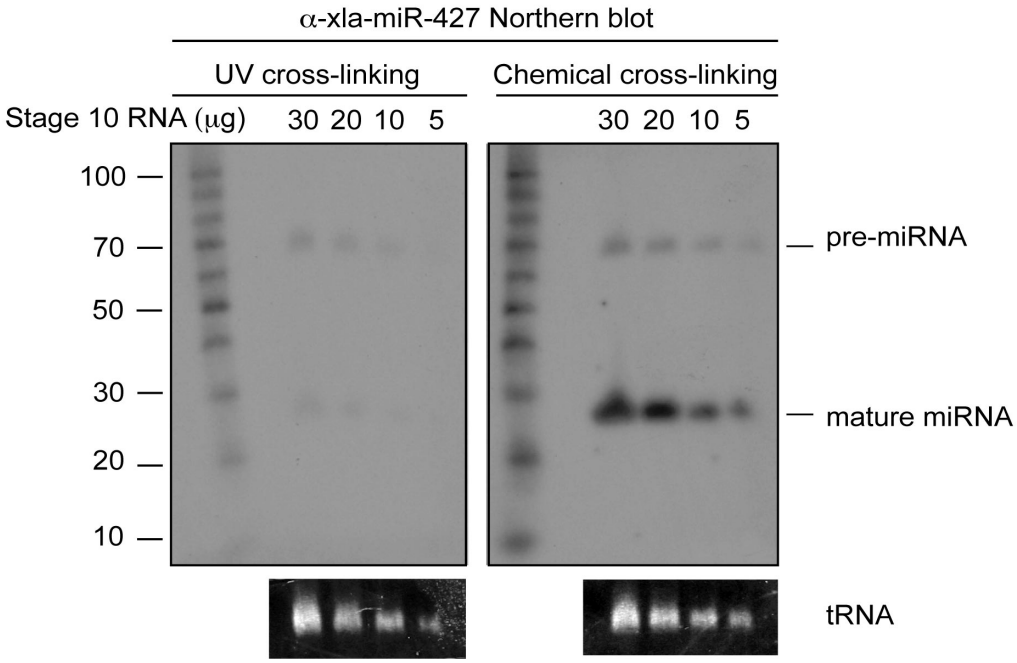


Figure 4.21 Chemical cross-linking improves sensitivity of small RNA detection by Northern Blot

Northern blots for xla-miR-427 in *Xenopus laevis* at stage 10 were cross-linked by ultra-violet light or 1-ethyl-3-3-dimethylaminopropyl carbodiimide (EDC) chemical cross-linking. 30, 20, 10 and 5 μg of *X. laevis* total RNA was probed. Ethidium bromide staining of tRNA serves as a loading control. RNA sizes (nt) were determined by comparison to radiolabelled decade markers (Ambion).

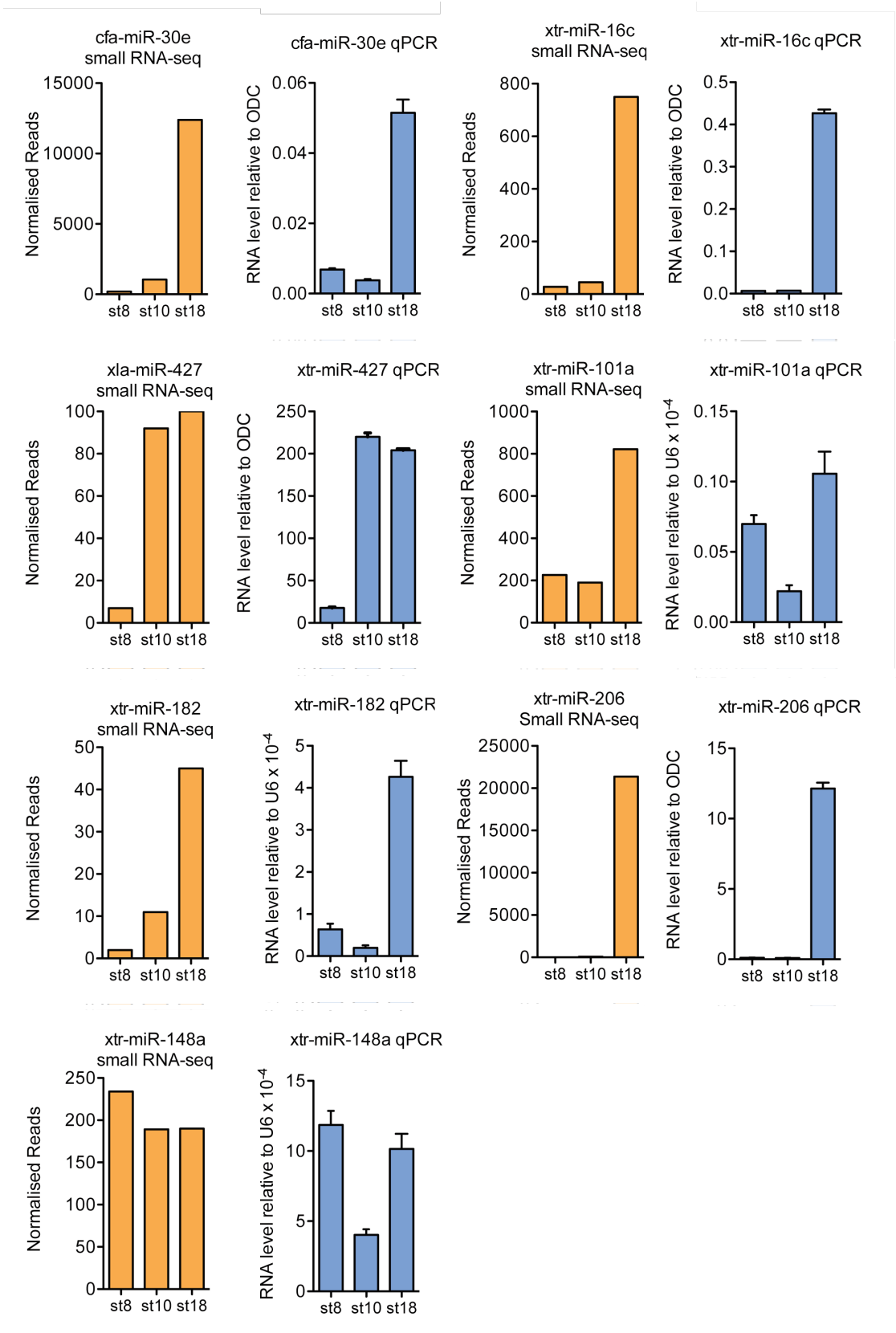


Figure 4.22 Successful qPCR validation of dynamically expressed microRNAs

MicroRNAs with qPCR profiles closely matching small RNA-seq data. See Figure 4.23 for a full legend.

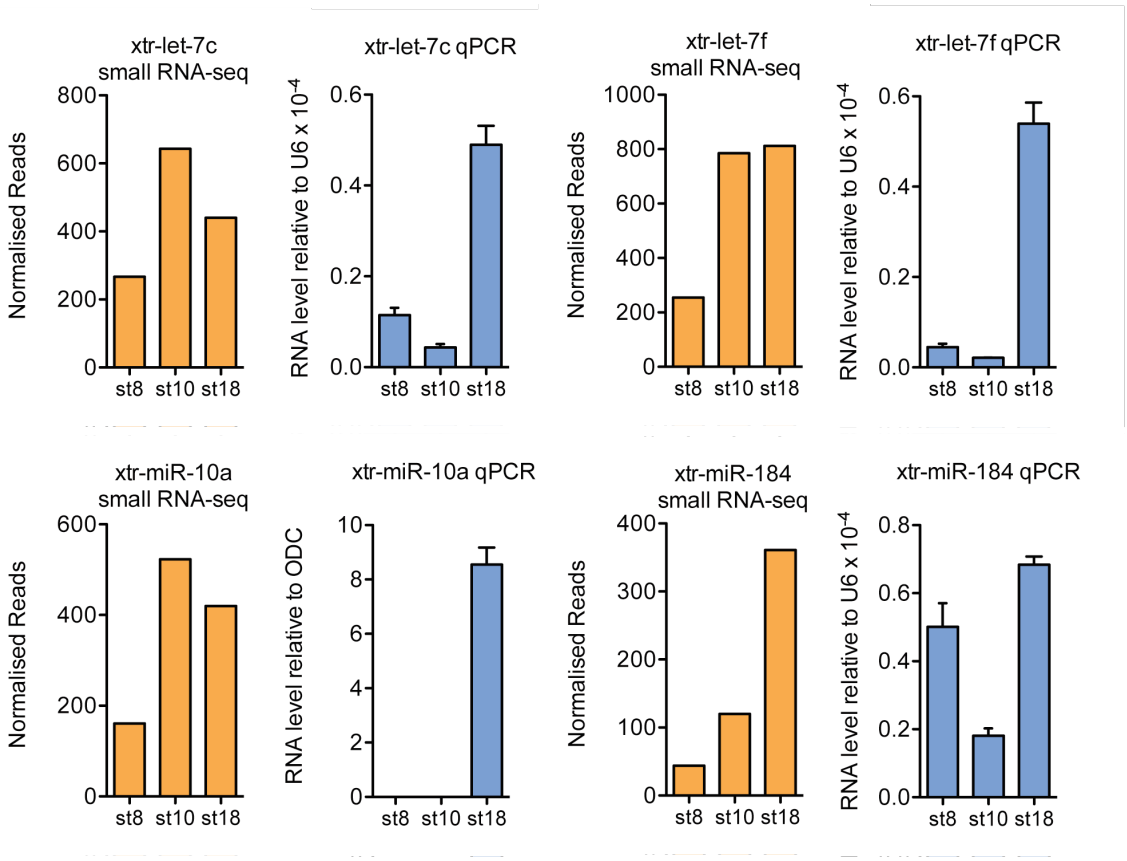


Figure 4.23 MicroRNA qPCR profiles of dynamic microRNAs with a mismatch in miRNA levels compared to small RNA-seq

X. tropicalis total RNA from whole embryos at stages 8, 10 and 18 was prepared for qPCR as in Figure 4.19 and microRNAs exhibiting dynamic expression patterns by Illumina small RNA sequencing were detected by qPCR and relative RNA levels to *ODC* or *U6* were calculated. Error bars are standard deviations from triplicate repeats. qPCR profiles (blue) were compared to the normalised microRNA read number (orange) obtained by small RNA sequencing in each library. Whereas the microRNA qPCR profiles shown in Figure 4.22 closely match the small RNA-seq read profiles, the qPCR profiles in this figure show lower microRNA levels at one stage than obtained in the small RNA-seq data. Note that in Figure 4.22, the xtr-miR-427 primer detects what is annotated as xla-miR-427 and xtr-miR-427 in miRbase.

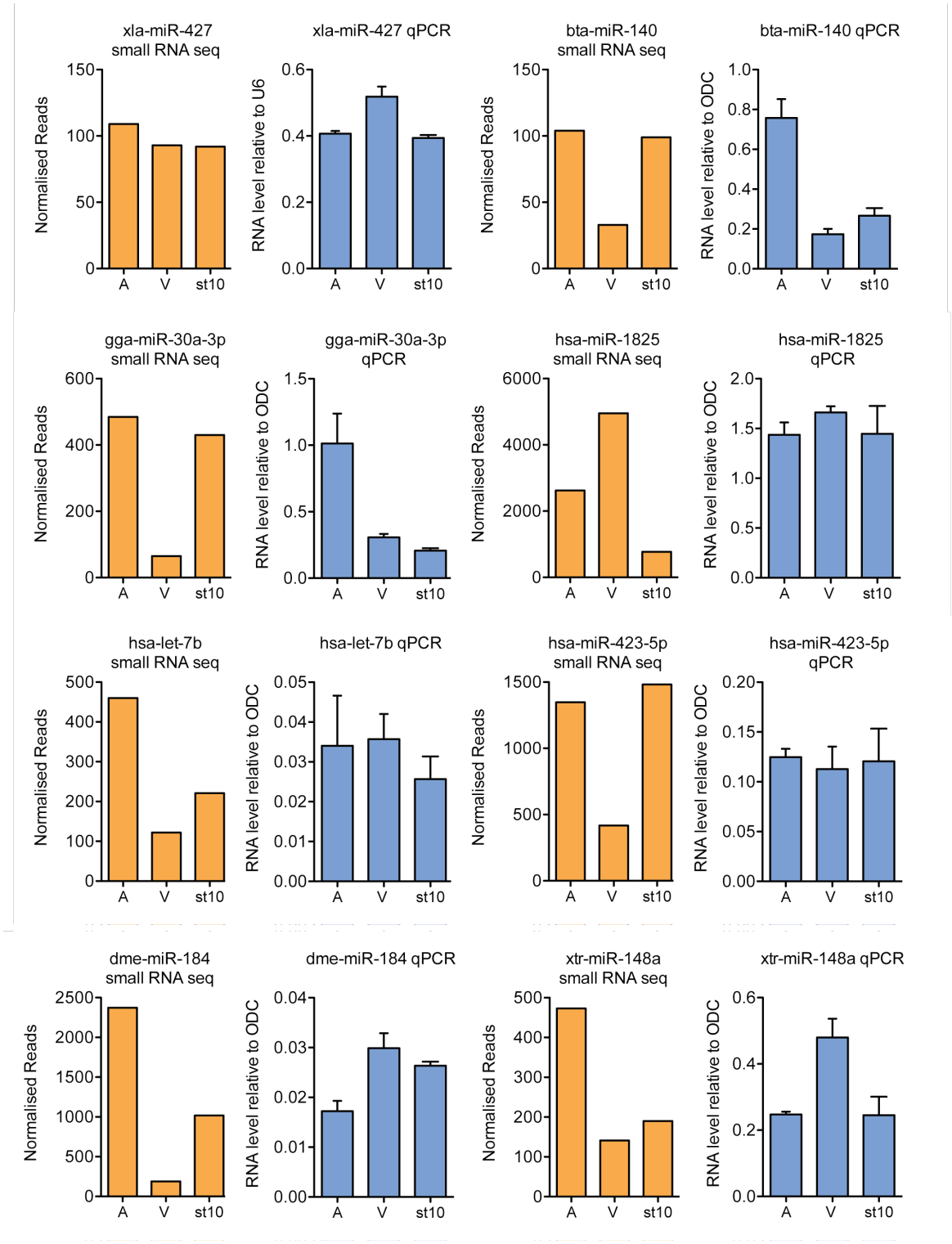


Figure 4.24 miRNA qPCR validation of spatially localised miRNAs

qPCR profiles of spatially localised miRNAs normalised to *ODC* or *U6* (blue) were compared to the miRNA normalised read number obtained by small RNA sequencing (orange) in the stage 10 animal pole (A), stage 10 vegetal pole (V) and stage 10 whole embryo (st10) small RNA libraries. Error bars are standard deviations from triplicate repeats.

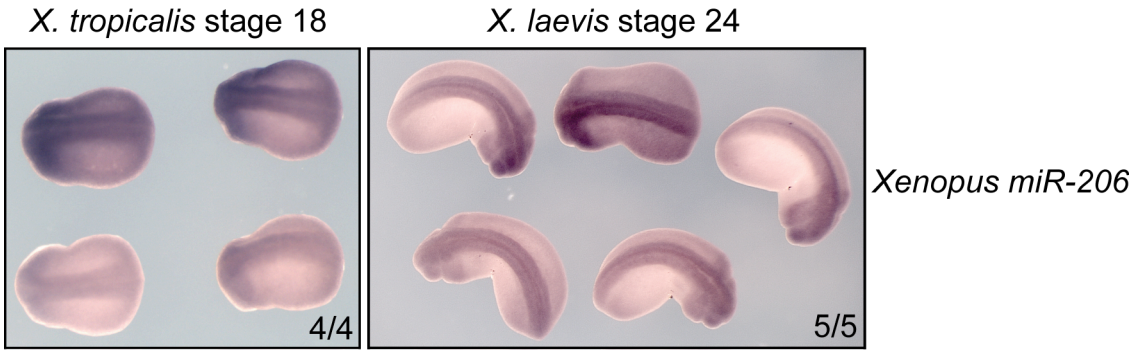


Figure 4.25 *Xenopus* miR-206 shows tissue-specific expression in muscle

Stage 18 *Xenopus tropicalis* and stage 24 *Xenopus laevis* embryos (shown here in dorsal view) were subjected to WISH using an xtr-miR-206 antisense LNA probe which was end-labelled with DIG. Probe hybridisation was detected with an anti-DIG antibody followed by staining in BM Purple. The number of embryos out of the total analysed with the staining pattern shown is indicated.

4.2.4.3 Piwi-interacting RNA validation

Experimental validation of piRNA expression by qPCR was only possible for the most abundant piRNA sequenced, PIR-32037. The small RNA read profile for PIR-32037 was consistent with the qPCR profile (Figure 4.26). All other piRNA qPCR primers tested gave rise to non-specific detection when qPCR dissociation curves were inspected. This is probably due to inefficient polyadenylation of the 3' end of piRNAs due to the presence of a methyl group on the 2' hydroxyl group of ribose (Kirino and Mourelatos, 2007). Northern blot detection of piRNAs in *X. tropicalis* was used to validate dynamic and spatially localised piRNAs. PIR-75450 was selected for experimental validation due to its stage-specific expression at stage 18 in the small RNA sequencing. Northern blot analysis with an antisense probe for PIR-75450 detected bands of approximately 30-32 and 70 nt (Figure 4.26). This could correspond to PIR-75450, which is 28 nt in the rat, and a biogenesis intermediate at 70 nt, as the mechanism of piRNA biogenesis is largely uncharacterised (Brennecke et al., 2007). By Northern blot, PIR-75450 is not expressed at stage 8 (fitting with zero reads in the small RNA seq at stage 8) and is weakly expressed at stage 10, increasing to a strong signal at stage 18. Overall, this expression profile fits with the small RNA seq results, except that low-level expression of PIR-75450 was not detected by small RNA seq at stage 10.

PIR-63460 shows stage-specific expression at stages 8 and 18 by small RNA seq, whereas Northern blot analysis showed bands of approximately 30 and 70 nt at all stages (Figure 4.26). The expected size of PIR-63460 is 32 nt and fewer than 100 small RNA seq reads were obtained, which may account for the weak signal on the Northern blot. PIR-75449 and PIR-72436 were highly abundant and vegetally localised according to the small RNA seq data with over 1000 vegetal reads at stage 10. Northern blot analysis in Figure 4.27 ratified the vegetal localisation and detected both piRNAs at stage 10, despite the lack of reads at stage 10 in small RNA-seq.

In short, the dynamic and spatial expression of the piRNA tags sequenced matched the experimental validation in terms of predicted abundance by small RNA seq, however the piRNA tags seemed to be slightly longer in *X. tropicalis* than the mammalian homologues used to annotate the piRNA tags. Importantly, the existence of

piRNAs later in development than the oocyte stage (see section 1.5.3) was experimentally confirmed. A 70 nt RNA species was detected in all piRNA Northern blots, which may be a biogenesis intermediate.

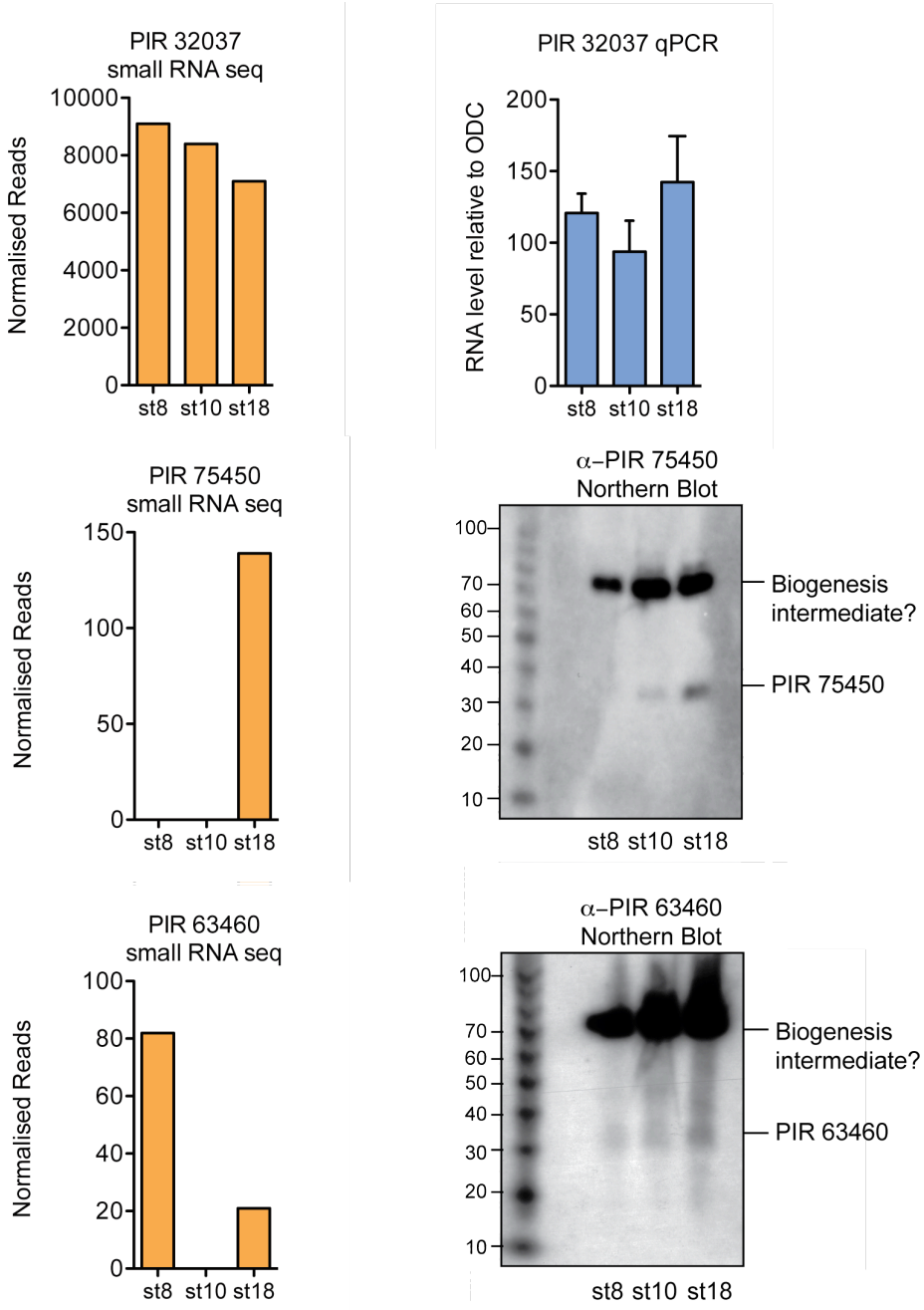


Figure 4.26 Dynamic piRNA validation

The number of normalised piRNA reads obtained by small RNA sequencing in *X. tropicalis* at stages 8, 10 and 18 (orange bars) is compared to qPCR (blue bars) or Northern blot expression at the same stages in *X. tropicalis* whole embryos for the piRNAs PIR 32037, PIR 75450 and PIR 63460. For the qPCR, piRNA levels were normalised to ODC and the error bars are standard deviations of triplicate repeats.

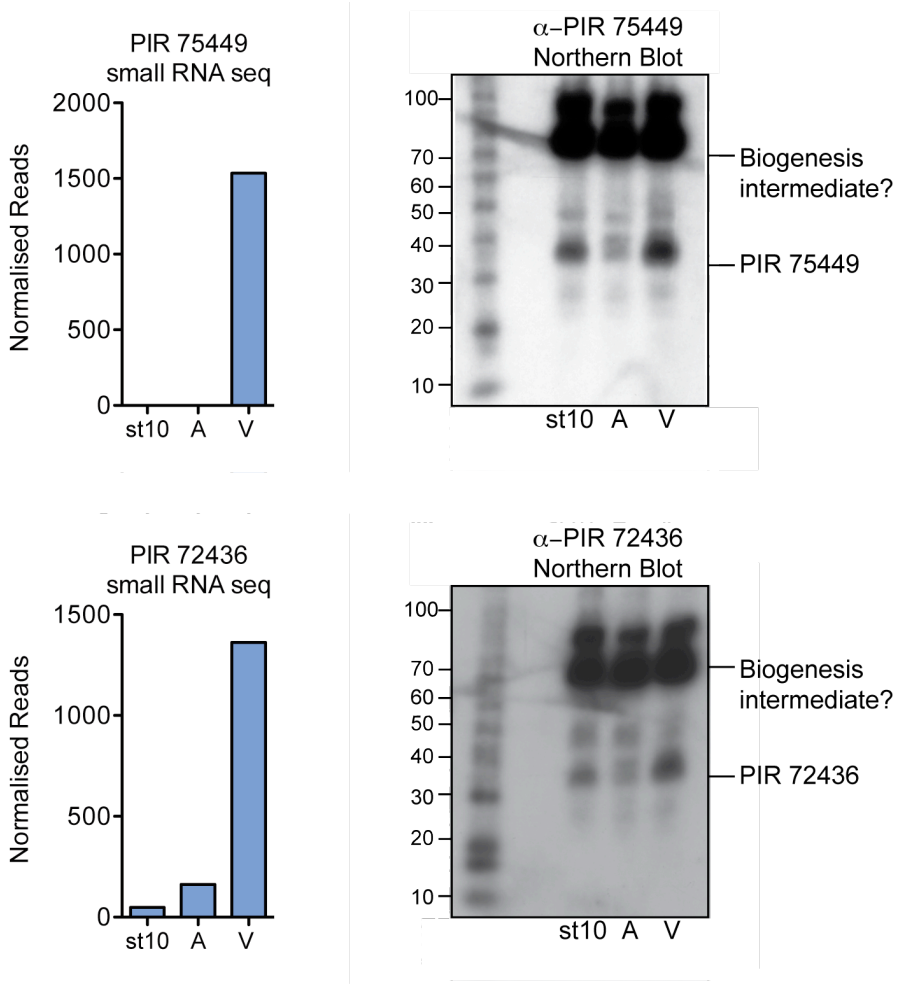


Figure 4.27 Spatial piRNA validation

The number of normalised piRNA reads obtained by small RNA sequencing in *X. tropicalis* in the stage 10 whole embryo (st10), stage 10 animal (A) and stage 10 vegetal (V) libraries (orange bars) is compared to Northern blot expression in *X. tropicalis* for the piRNAs PIR 75449 and PIR 72436 which are vegetally localised in the small RNA seq libraries at stage 10. Sizes of small RNAs were determined by comparison to decade markers (Ambion).

4.3 Discussion

4.3.1 Summary of results

- Over 14 million small RNAs are present in the early *Xenopus tropicalis* embryo.
- MicroRNAs are less than 1.6 % of the small RNAome in early development, and piRNAs are less than 0.9 %.
- Over 95 % of the small RNAs in early development are as yet uncharacterised.
- 169 microRNAs and 64 piRNAs were identified in *X. tropicalis* early development by stringent comparison to known databases, of which 56 microRNAs were already known in *Xenopus*. A further 23 miRNAs were identified by re-sequencing the libraries, including 3 known *Xenopus* miRNAs.
- MicroRNAs present at stage 8 and stage 10 by small RNA-seq were analysed for seed matches in the minimal 276-520 nt region of *XCR1α* 3'UTR responsible for spatial regulation (see Figure 3.27).
- MicroRNAs increase in abundance and diversity between blastula and late neurula stages of *X. tropicalis* development and are more abundant in the animal than the vegetal pole at gastrulation.
- Piwi-interacting RNAs are present in the developing embryo and with stable expression from blastula to gastrulation and no general trend for localisation to the animal or vegetal pole at gastrulation.
- Specific microRNAs and piRNAs are dynamically and spatially localised in early development.
- Thorough experimental validation, a technical sequencing repeat and stage 10 small RNA library biological repeats confirm the quality of the small RNA sequencing dataset.

This study is the first to characterise the small RNAome of early *X. tropicalis* development, and one of the first studies comparing stages of early vertebrate development. Previously, very little was known about microRNA expression in early vertebrate development. Most *Xenopus tropicalis* microRNAs listed in miRbase had been identified in adult tissues (Tang and Maxwell, 2008) and 26 mature *Xenopus laevis* miRNA sequences were identified at stages from oocyte to tadpole (Watanabe et al., 2005). The small RNA profile of the *X. tropicalis* oocyte and adult liver and skin tissues had been characterised by small RNA sequencing (Armisen et al., 2009). In total, I have identified 192 microRNAs (169 from the first CRI sequencing run and and additional 23 from the technical repeat) that are present in at least one stage of blastula to late neurula development of *X. tropicalis*. Of these miRNAs, 133 were identified by comparison to microRNAs identified in other (non-*Xenopus*) species. This is strong evidence for conservation of microRNAs across species. In total, I sequenced 59 known *Xenopus* microRNAs in the early development small RNA libraries, out of a possible 182 known *Xenopus* microRNAs listed in miRbase. Therefore about 32 % of the known *Xenopus* microRNAs are present in early development. This fits with the result that microRNA expression is highly dynamic between stages 10 and 18, with microRNA abundance and sequence diversity increasing dramatically with time in development. This agrees with a previous microarray study of microRNA expression in early zebrafish development that showed a general trend of increasing microRNA abundance with hours post-fertilisation (Wienholds et al., 2005).

Moreover, microRNAs were less than 1.6 % of the small RNAome in early development. This fits with previous work comparing the microRNA profile of the *X. tropicalis* oocyte and adult tissues, where 34 and 114 known *X. tropicalis* microRNAs were sequenced at read numbers greater than 10 respectively (Armisen et al., 2009). Interestingly, xtr-miR-10b, xtr-miR-101 and xtr-miR-148a were the most frequently sequenced microRNAs in the oocyte, which correlates with their expression at stage 8 in the small RNA libraries I sequenced (Figure 4.23). Notably, both xtr-miR-101a and xtr-miR-148a decrease at stage 10 compared to stage 8, and increase again at stage 18. This suggests that maternally abundant miRNAs do undergo some regulation at the mid-blastula transition when maternal RNAs are degraded with the onset of zygotic transcription.

MicroRNA expression increasing between stages 10 and 18 could be due to the increase in the number of tissues from three at gastrulation (endoderm, mesoderm and ectoderm) to the onset of multiple tissue specification at stage 18, for example neural crest and somites are specified post-gastrulation. MicroRNAs are thought to fine-tune gene expression; for example, muscle specific miRNAs including miR-206 are upregulated at the onset of skeletal muscle development in human cell lines by myogenic transcription factors Myogenin and MyoD (Rao et al., 2006). Tissue-specific expression of microRNAs has been documented at later developmental stages in *X. tropicalis* in organs such as the eye, and brain and in the nervous system (Walker and Harland, 2008).

The five most frequently sequenced microRNAs at stage 18, xtr-miR-206, cfa-miR-30e, mmu-miR-10b, rno-miR-10a-3p and xtr-miR-130b comprise 66 % of total microRNA reads. This suggests that the stage 18 microRNA profile is dominated by a small number of highly abundant microRNAs, and it would be exciting to further explore the functional role of these microRNAs later in development with loss of function studies. Moreover, I found that microRNAs were generally more abundant in the animal than the vegetal pole at stage 10 and validated the animally localised microRNAs bta-miR-140 and gga-miR-30a-5p. MicroRNAs may play a role in fine-tuning the expression of genes expressed in the prospective ectoderm, such as genes involved in specification of tissues derived from ectoderm such as the epidermis and nervous system. This would explain why tissue-specific expression of microRNAs is often observed in the nervous system of *X. tropicalis* (Walker and Harland, 2008) and zebrafish (Kapsimali et al., 2007).

Many microRNAs sequenced in this study matched to microRNA homologues in multiple species; for example, the mouse, *X. tropicalis*, human and horse homologues of miR-182 were all sequenced (see Figure 4.8). By definition, the sequences of these microRNA homologues are distinct, as they correspond to different tags. Their sequences are shown in Table 4.9. The presence of miR-182 homologues from multiple species in the *X. tropicalis* small RNA libraries could be due to several reasons. The miR-182 sequences differ by one nucleotide at the 5' end (mmu-miR-182) or 2-3 nucleotides at the 3' end. This could be an artefact of trimming the raw reads to remove

adapter sequences in the case of shortening eca-miR-182 by one G base to give the hsa-miR-182 sequence, or genuine editing of the end nucleotides of the mature miRNAs (Kawahara et al., 2007b; Seitz et al., 2008). Adenosine deaminase-dependent microRNA editing is more often observed in the primary miRNA transcript (Kim et al., 2010) (Kawahara et al., 2007a) and is limited to conversion of adenosine to inosine, which base-pairs with cytosine instead of uracil. This means that this type of miRNA editing can only account for U to C alterations in miRNA sequence. However, microRNAs have been shown to display genuine sequence variability by addition or deletion of 1-2 nt at the 3' end in *Drosophila, C. elegans* and mouse (Seitz et al., 2008). Therefore the similarity of the *X. tropicalis*, human and horse miR-182 sequences sequenced in *X. tropicalis* could be explained by inherent 3' variability from Drosha/Dicer processing, which yields several microRNA tags that are identical in sequence to a miR-182 homologue from another species. Another possibility is the existence of a miR-182 microRNA family in *X. tropicalis*, which is plausible given that the mmu-miR-182 sequence cannot be formed by addition/deletion of 1 or 2 nucleotides from the *X. tropicalis*, human and horse sequences (see Table 4.9).

MicroRNA homologue sequenced in *X. tropicalis*	**miRNA sequence**
mmu-miR-182 (mouse)	5' UUGGCAAUGGUAGAACUCACACCG
xtr-miR-182 (*X. tropicalis*)	5' UUUGGCAAUGGUAGAACUCACA
hsa-miR-182 (human)	5' UUUGGCAAUGGUAGAACUCACACU
eca-miR-182 (horse)	5' UUUGGCAAUGGUAGAACUCACACUG

Table 4.9 miR-182 homologues sequenced in *X. tropicalis*

11 let-7 family microRNAs sharing the seed sequence 5' GAGGUA were sequenced in the *X. tropicalis* small RNA libraries (Figure 4.11), including four known *X. tropicalis* let-7 miRNAs, human, *Drosophila*, bull, opossum and mouse let-7 homologues. Their sequences are listed in Table 4.10 and show that single nucleotide

differences in the middle of the miRNA sequence and differences in the end nucleotides are present in a microRNA family.

let-7 family microRNA homologue sequenced in *X. tropicalis*	**miRNA sequence**
xtr-let-7a (*X. tropicalis*)	5’ UGAGGUAGUAGGUUGUAUAGUU
xtr-let-7c (*X. tropicalis*)	5’ UGAGGUAGUAGGUUGUAUGGUU
xtr-let-7f (*X. tropicalis*)	5’ UGAGGUAGUAGAUUGUAUAGUU
hsa-let-7b (human)	5’ UGAGGUAGUAGGUUGUGUGGUU
hsa-let-7d (human)	5’ AGAGGUAGUAGGUUGCAUAGUU
hsa-let-7e (human)	5’ UGAGGUAGGAGGUUGUAUAGUU
hsa-let-7f (human)	5’ UGAGGUAGUAGAUUGUAUAGUU
dme-let-7 (*Drosophila*)	5’ UGAGGUAGUAGGUUGUAUAGU
bta-let-7e (bull)	5’ UGAGGUAGGAGGUUGUAUAGU
mdo-let-7d (opossum)	5’ AGAGGUAGUAGGUUGCAUAGU
mmu-let-7g (mouse)	5’ UGAGGUAGUAGUUUGUACAGUU

Table 4.10 let-7 family microRNAs sequenced in *X. tropicalis*

The biological roles of miRNAs in early *Xenopus* development have only been characterised for a handful of miRNAs. miR-427 modulates Nodal signalling by repressing the expression of Nodal ligands (Xnr 5 and Xnr6) and the Nodal inhibitor Lefty (Rosa et al., 2009) and I have shown that miR-427 negatively regulates XCR1 expression in the animal cap, as discussed in section 3.3.2. Modulation of Nodal signalling at the receptor level by *Xenopus* miR-15 and miR-16 has also been documented (Martello et al., 2007). To determine whether any of the microRNAs I sequenced in *X. tropicalis* modulated Nodal signalling by targeting XCR1, I screened the list of microRNAs I sequenced at stages 8 and 10 for exact seed matches to the region of the *XCR1α* 3’UTR I had shown was responsible for spatial regulation (276-520 nt). xtr-miR-107, cel-miR-44, dme-miR-279, gga-miR-1810, bta-miR-140 and hsa-

miR-140-3p passed this criteria, however it is worth noting that one mismatch was allowed when comparing small RNA-seq reads to the miRbase miRNA sequence and so cel-miR-44, gga-miR-1810 and hsa-miR-140-3p are not listed in the heatmaps in Figures 4.8-4.11 which use a zero mismatch filter. Other biological roles of miRNAs in early *Xenopus* development involve clearance of maternal mRNA by miR-427 (Lund et al., 2009). *X. tropicalis* miR-206 probably functions in muscle development given its role in human muscle development (Rao et al., 2006), expression in somites in *X. tropicalis* and *X. laevis* (Figure 4.25) and identical mature sequence to human miR-206. As very few microRNAs have been assigned a function in early development, the function of the 133 microRNAs I sequenced that are new to *Xenopus* could be explored further by knockdown studies, as well as using the small RNA-seq data to investigate the mechanism of XCR1 spatial regulation.

Previous studies of piwi-interacting RNAs have focussed on their role preserving genome integrity in the germline against transposon integration during germline development in *Drosophila*, zebrafish and *C. elegans* (Brennecke et al., 2007; Houwing et al., 2007; Ruby et al., 2006). In *Xenopus*, Piwi family proteins have been found in the oocyte, but not in somatic cells (Wilczynska et al., 2009). Similarly, piRNAs were exclusively expressed in *X. tropicalis* oocytes and not in adult tissues in a previous small RNA sequencing study (Armisen et al., 2009).

In this small RNA sequencing study, I found 64 piRNAs in the developing *X. tropicalis* embryo. The lack of piRNAs in adult tissues in the previous *X. tropicalis* small RNA sequencing study may be due to relatively poor sequencing depth: approximately 6 million small RNA reads were obtained per library, in contrast to over 14 million reads per library in my study. However, the piRNAs analysed by Armisen *et al*. had been more rigorously experimentally validated than in my work by confirming an interaction of the piRNA with PIWI and the presence of a 2'*O*-methyl modification at the 3' end of the piRNA (Armisen et al., 2009). piRNAs have also been found in somatic tissues of *C. elegans* (21U-RNAs), *Drosophila*, mouse and the Rhesus macaque (Kato et al., 2009; Yan et al., 2011), which fits with my observation of piRNA sequences in *X. tropicalis* early development and suggests that piRNAs may be important outside of germline development. This will be discussed further in Chapter 6.

Previous small RNA-seq studies have focused largely on microRNA profiling of human disease, for example, leukemia development in a mouse model (Kuchenbauer et al., 2008), or comparing microRNA profiles of normal tissue and tumour samples (Farazi et al., 2011), comparing microRNA expression during tumour progression to identify biomarkers (Martens-Uzunova et al., 2011) and miRNA profiling of stem cells (Morin et al., 2008). MicroRNA profiling by small RNA-seq has been used for novel microRNA discovery in organisms ranging from unicellular algae to silkworms to the ancient vertebrate amphioxus (Chen et al., 2009; Liu et al., 2010; Molnar et al., 2007). However, the only other study of small RNAs in early vertebrate development apart from this thesis to date is in the chicken embryo (Glazov et al., 2008). Small RNA-sequencing of days 5, 7 and 9 of chicken embryo development revealed a statistically significant increase in the expression of 7 miRNAs and a decrease in the expression of 22 miRNAs, which the authors speculate could be due to a restriction in microRNA expression domains as development progresses, which is contrary to my observations and could be due to the relatively late time points in chicken development sequenced.

Small RNA-seq analysis of stages in *C. elegans* development ranging from embryo to young adult showed that the proportion of microRNA reads was 33 % in the *C. elegans* embryo, increasing to 61 % the first larval stage (L1), 69 % in the fourth larval stage (L4) and was 65 % in the young adult worm (Kato et al., 2009). This agrees with my conclusion that microRNA expression increases with time in development (Table 4.4 and Figure 4.17A). However, it is striking that a much larger proportion of the small RNA reads were identified as microRNAs in *C. elegans*, compared to less than 1.6 % of the early *X. tropicalis* embryo in my study. MicroRNA expression in *C. elegans* was also found to be dynamic: 16 % of the sequenced miRNAs changed in expression at least 10-fold between embryo and young adult stage. This agrees with the dynamic expression of miRNAs I observed in Figures 4.8 to 4.11. Interestingly, *C. elegans* piRNAs, known as 21U-RNAs, increased from 0.7 % of reads in the third larval stage (L3) to a maximum of 6.8 % of small RNA reads in the young adult stage (Kato et al., 2009). This suggests that piRNAs are more abundant and dynamic in *C. elegans* development compared to early *X. tropicalis* development where piRNAs are less than 0.9 % of small RNA reads, and supports the conclusion from my work and other studies that piRNAs are present in somatic tissues.

Although I sequenced many dynamic and spatially localised *X. tropicalis* microRNAs and piRNAs in this study, microRNAs were less than 1.6 % and piRNAs less than 0.9 % of the small RNA reads sequenced. In total, less than 5 % of reads were annotated by comparison to known databases. The remaining 95 % of unannotated reads align to the genome and are genuine small RNAs rather than mRNA degradation products, as mRNA degradation products lack the necessary 5' phosphate and 3' hydroxyl groups required for small RNA library construction (Hafner et al., 2008). This raises the exciting possibility that these unannotated reads may represent a new class of small RNAs that mediate a novel mode of gene regulation. The annotation and function of the unknown reads will be explored in Chapter 5.

Chapter 5. Analysis of unannotated small RNAs

5.1 Introduction

The aim of the work presented in this chapter was to characterise the 95 % to 98 % of the *X. tropicalis* small RNAome that could not be annotated by comparison to databases of known small RNAs, but aligns perfectly to the genome. These uncharacterised small RNAs potentially represent novel regulatory small RNAs. Firstly, the unannotated reads are fed into a novel microRNA prediction pipeline to yield candidate microRNAs. One candidate microRNA, provisionally named miR-F, is characterised in detail. Secondly, I analyse the results of a genome-wide annotation of the small RNAs based on where they align in the genome and proximity to other small RNAs in clusters. A possible regulatory role for an exciting class of novel small RNAs that cluster in introns is discussed.

5.2 Results

5.2.1 Novel microRNA discovery

In order to reveal novel microRNAs within the unannotated small RNAs, I designed a *de novo* microRNA prediction pipeline, which was constructed and operated by Stuart Horswell in the Cancer Research UK Bioinformatics and Biostatistics group. All unannotated tags from all five small RNA libraries of read number greater than 10 reads and a length of between 22 and 23 nt were selected for novel microRNA analysis. This eliminated noise from low abundance tags and reads outside the size range for miRNA. The reads were aligned to the *X. tropicalis* genome with up to two mismatches permitted and contigs of overlapping reads (tag-contigs) were assembled (Figure 5.1) (Jung et al., 2010). Bases are assigned a weighted depth by calculating the number of overlapping reads per base. The genomic sequence of tag-contigs with a contiguous depth greater than 1 and length greater than 22 nt was retained and all other sequences were discarded. The genomic sequence of the tag-contig was extended by 10 nucleotides 5’ and 60 nucleotides 3’ or vice versa to form a candidate miRNA precursor sequence (pre-miRNA). The putative pre-miRNA sequence was folded into a secondary structure by minimising the free energy of folding using Mfold (Zuker, 2003). Resulting

secondary structures were screened against the criteria for miRNA precursors, such as a hairpin secondary structure without large bulges, the mature miRNA sequence contained within one arm of the hairpin and at least 16 bp of the mature miRNA sequence base pairing with the candidate microRNA-star sequence (Ambros et al., 2003a). Six candidate microRNAs passed these criteria provisionally named miR-A, B, C, D, E and F (Figure 5.2).

I then attempted to experimentally confirm the expression of the novel candidate microRNAs at stages 8, 10 and 18 in *X. tropicalis* development by miRNA-qPCR. miR-A, miR-C and miR-F were experimentally validated (Figure 5.3). However, non-specific amplification was observed for miR-B, miR-D and miR-E as multiple non-symmetrical peaks were observed in the qPCR dissociation curves. This is a successful prediction rate for the novel microRNA prediction pipeline of 50 %. Interestingly, miR-F was predicted to exist in the stage 18 library and this was experimentally ratified by qPCR. miR-A and miR-C were predicted to exist in the stage 8 and stage 10 animal libraries respectively, but were found to be expressed at stage 8, 10 and 18 by qPCR. Sequences and genomic locations for experimentally validated miRNAs are presented in Table 5.1. MicroRNAs are officially named according to sequence homology with other miRNAs by the miRbase miRNA naming service only after a manuscript is accepted for publication; hence provisional miRNA names are used throughout this chapter. In short, three completely novel microRNAs in early *X. tropicalis* development were confidently identified from the unannotated reads obtained in small RNA sequencing; xtr-miR-A, xtr-miR-C and xtr-miR-F.

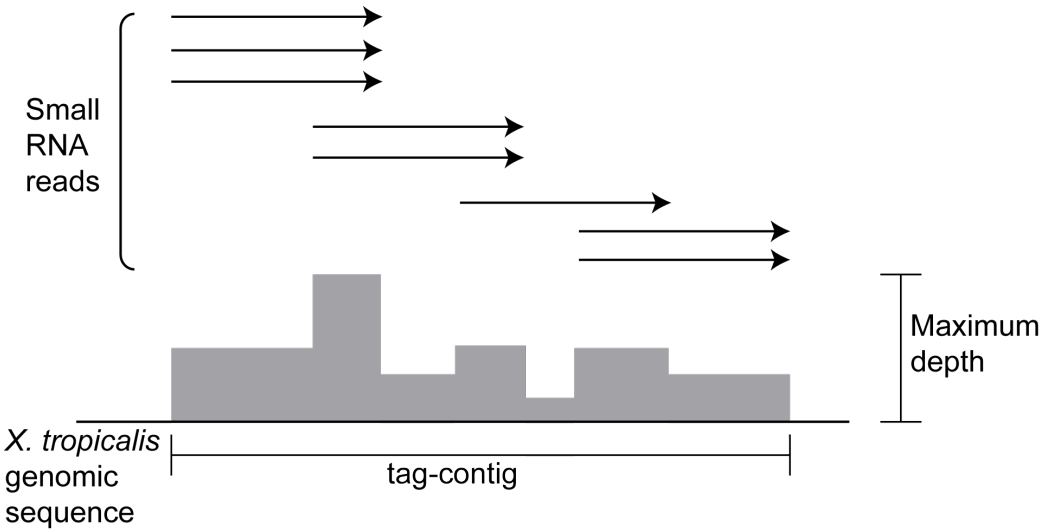


Figure 5.1 Tag-contig assembly from overlapping small RNA reads

Tags with read numbers greater than 10 and length of 22-23 nt were aligned to the genome with 2 mismatches from all five small RNA libraries. Overlapping small RNA reads form tag-contigs. Bases in the tag-contig are assigned a weighted depth by calculating the number of overlapping reads per base. Only tag-contigs of a maximum depth greater than 1 and length greater than 22 bases were retained for the next step of novel microRNA prediction.

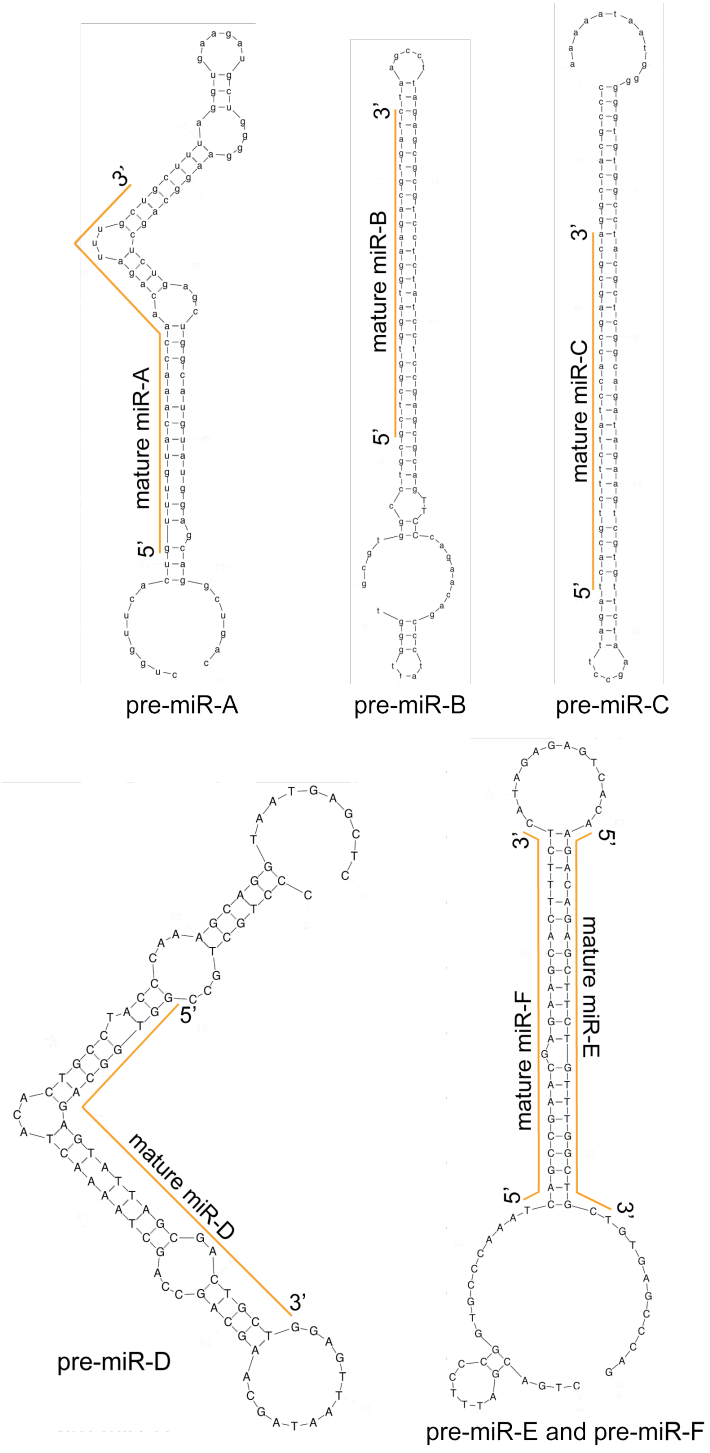


Figure 5.2 Six candidate miRNAs passed hairpin folding criteria

The genomic sequence of tag-contigs was extracted, extended and folded into secondary structures based on minimising the free energy of folding using Mfold. Six candidate microRNAs formed secondary structures obeying the rules for potential microRNA precursors (pre-miRNAs), provisionally named miR-A-F. Note that miR-E and miR-F are predicted to be a miRNA* and mature miRNA pair respectively. A yellow line indicates mature miRNA sequences within the hairpin precursor.

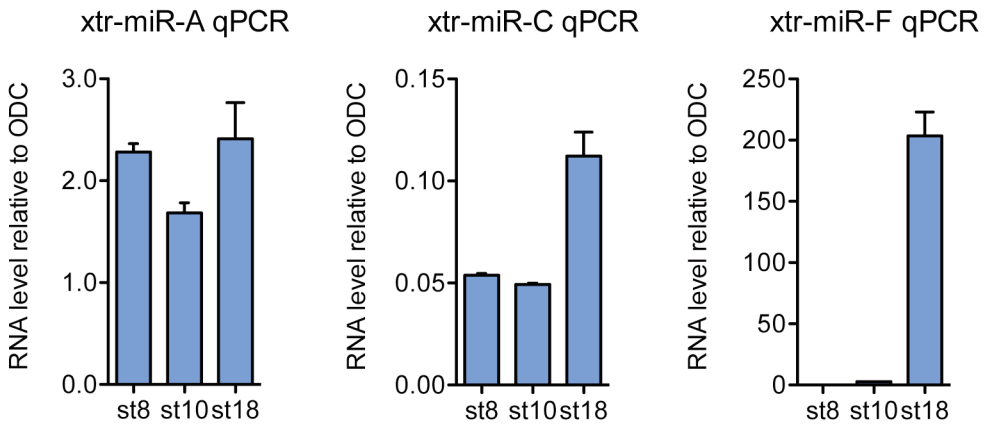


Figure 5.3 Experimental validation of novel microRNAs

X. tropicalis total RNA from whole embryos at stages 8, 10 and 18 was prepared for qPCR as in Figure 4.8 and the candidate novel microRNAs xtr-miR-A, xtr-miR-C and xtr-miR-F were detectable as a single PCR product by qPCR and relative RNA levels to ODC were calculated. Error bars are standard deviations from triplicate repeats. xtr-miR-B, xtr-miR-C and xtr-miR-E were not detectable by qPCR.

Provisional miRNA name	Sequence of mature miRNA	Genomic location	Small RNA library predicted in
xtr-miR-A	5’agcaaatctgttggtttgtacaaac 3’	scaffold 1337	stage 8
xtr-miR-C	5’tgcgctcggtggatagaagacgtga 3’	scaffold 212	stage 10 animal
xtr-miR-F	5’gagaaagtgcttctcgttcggctga 3’	scaffold 1310	stage 18

Table 5.1 Experimentally validated novel microRNAs

The sequences of the mature novel microRNAs that were confidently identified are shown with their approximate genomic locations and the small RNA library that they were predicted to exist in for comparison with the experimentally determined expression in Figure 5.3.

5.2.2 Novel microRNA function and conservation

xtr-miR-F was also experimentally validated in *X. tropicalis* by Northern blot (Figure 5.4), whereas no signal was detectable for xtr-miR-A and xtr-miR-C, possibly due to their low abundance. xtr-miR-F was therefore studied in more detail. Mature xtr-miR-F is expressed in *X. tropicalis* from stage 18 to stage 40 (Figure 5.4), however the precursor miRNA (pre-xtr-miR-F) is present at stages 8 and 10, which indicates that the processing of mature miR-F is developmentally regulated. xtr-miR-F is also conserved in *Xenopus laevis*, as both pre-miRNA and mature miRNA bands are present when total *X. laevis* RNA was probed with an antisense probe to the xtr-miR-F sequence (Figure 5.5A). The expression profile of xla-miR-F was very similar to that of xtr-miR-F, with mature miRNA expression beginning at stage 18 and persisting until tadpole stage. The onset of mature xla-miR-F expression was refined by probing *X. laevis* RNA from stages 10 through to 20 for xla-miR-F. Mature xla-miR-F is processed out from the hairpin precursor at stage 12, the beginning of neurulation, and reaches maximal levels at stage 15 (Figure 5.5B). This stage-specific expression suggests that miR-F may play a role in regulation of gene expression during neurulation, as mature miR-F levels decline post stage 18 in *X. tropicalis* and after stage 26 in *X. laevis*.

In order to assess whether miR-F was dependent on the microRNA-processing enzyme Dicer for its expression, Dicer was knocked down by morpholino injection in *X. tropicalis* and in *X. laevis*. Knockdown of Dicer would be expected to decrease the level of Dicer-dependent mature miRNAs and increase the pre-miRNA levels. Figure 5.6A shows that a splice-blocking morpholino designed to inhibit splicing out of intron 1 from between exons 1 and 2 of *X. tropicalis* Dicer mRNA is effective at stage 18. Insertion of intron 1 introduces a premature stop codon in the mRNA sequence to produce a functionally inactive truncated protein. When xtr-miR-F levels are assessed in a dicer morphant background, there is a modest decrease in mature miR-F levels and increase in pre-miR-F (Figure 5.6B). When the ratio of pre-miRNA to mature miRNA is quantified, there is a 10.2 % increase in the pre-miRNA: mature miRNA ratio when Dicer morpholino is injected compared to control morpholino. Intriguingly, when the expression profile of Dicer mRNA in *X. tropicalis* development is assessed by qPCR, developmental regulation of Dicer expression is observed. Dicer mRNA increases dramatically from stage 10 to stage 18, in a similar manner to global microRNA

expression (Figure 4.6A). Previously, *Dicer* mRNA expression had been assessed by EST profiles from cDNA libraries (Bowes et al., 2010). According to this approximate gene expression profile of *Dicer*, mRNA is present from the neurula to adult stages but had not been identified in ESTs from libraries at egg or gastrula stage.

A translation blocking *X. laevis* Dicer morpholino was shown to be highly effective in Figure 3.21A. A small decrease in mature xla-miR-F expression at stage 38 was observed in *X. laevis* dicer morphant embryos compared to control embryos (Figure 5.7A). Dicer-dependency of xla-miR-427 was also analysed by Northern blot (Figure 5.7B). A decrease in mature xla-miR-427 was observed in Dicer morphants at stages 26 and 36 when no pre-miRNA is detectable. This suggests that Dicer knockdown may only affect mature miRNA levels at later stages, perhaps because some mature miRNAs are relatively stable; for example miR-223 in mammalian cells decays slowly over days when its biogenesis is terminated (Baccarini et al., 2011). In short, miR-F appears to be partially Dicer-dependent, as a modest increase in pre-miRNA and a reduction in mature xtr-miR-F is observed in *X. tropicalis* Dicer morphants at stage 18, and xla-miR-F expression is decreased at later stages in *X. laevis* Dicer morphants. It is also possible that miR-F could be processed by another enzyme (see section 5.3 for further discussion).

To explore the functional role of xtr-miR-F and xla-miR-F, I knocked down miR-F in both frog species using the same morpholino. Morpholino inhibition of xtr-miR-F in *X. tropicalis* led to defective axis elongation and posterior development (7/22 embryos), however this phenotype was not fully penetrant as 8/22 embryos were relatively normal albeit slightly delayed in development compared to control embryos (Figure 5.8). A small proportion of embryos were normal but curved (2/22 embryos, data not shown) and 5/22 embryos died between stage 18 and stage 38. Inhibition of xla-miR-F in *Xenopus laevis* led to a less penetrant axial/posterior development phenotype (2/44 embryos). Morphant embryos also displayed reduced eye size and fewer melanocytes (10/44 embryos). Neither of these phenotypes was fully penetrant, hence 29/44 morphant embryos were normal compared to control morpholino-injected embryos. In summary, miR-F leads to defective axial and posterior development to some extent in

both frog species, with an additional role in eye development and pigmentation in *Xenopus laevis*.

Since miR-F was conserved between frog species, I explored whether miR-F was also conserved in mammalian species. miR-F was found to be present in human keratinocyte cells (HaCats (Boukamp et al., 1988)), human lymphoblastoid cells (Levy et al., 1968), human breast carcinoma cells (MDA-MB-231 (Cailleau et al., 1974)) and human embryonic kidney cells (293Ts (Graham et al., 1977)) when these cell lines were tested for miR-F expression by qPCR (Figure 5.10). Excitingly, miR-F levels were higher in the normal cell line models (HaCat and lymphoblastoids) compared to the breast cancer cell line MDA-MB-231 and 293T cells which are transformed, which raises the possibility of human miR-F downregulation in cancer and a role as a tumour suppressor.

5.2.3 Novel microRNA target prediction

I initiated an unbiased approach to find miR-F target genes by preparing total RNA from control and xtr-miR-F morpholino-injected stage 18 *X. tropicalis* embryos for transcriptome profiling by genome-wide mRNA sequencing. Since microRNAs are thought to lead to degradation of their target mRNAs (Baek et al., 2008), miR-F target genes would be expected to be upregulated in stage 18 xtr-miR-F morphant embryos. A similar approach to find the targets of the let-7 microRNA in *C. elegans* was successful using a similar approach: let-7 was overexpressed and microarray analysis revealed downregulated candidate target genes (Johnson et al., 2007). Unfortunately, the coverage of the genome was very poor for the miR-F knockdown library. Of over 23 million raw mRNA-seq reads, only 0.88 % aligned to the *X. tropicalis* genome with zero mismatches. The quality of the mRNA seq data was too low to merit analysis and the mRNA libraries needing remaking in this case. However, the observation that miR-F is conserved in mammalian cells means that knockdown of miR-F and mRNA-seq can be performed in cell culture, thereby avoiding the possible indirect effects of miRNA knockdown caused by multiple tissue interactions in the embryo. Another approach could be to look for exact miR-F seed matches in the 3'UTRs of *X. tropicalis* or human genes; however, this is likely to yield an unworkably high number of candidate target genes.

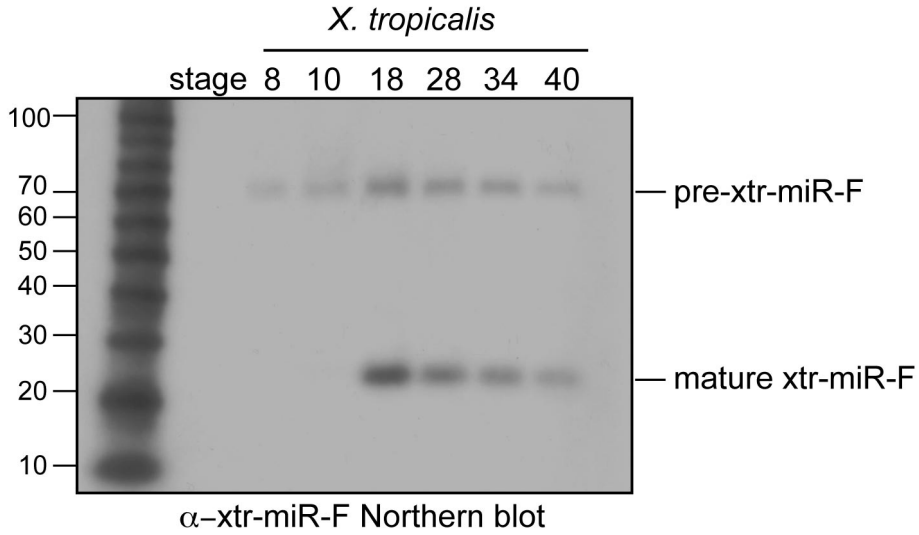


Figure 5.4 xtr-miR-F is expressed from late neurula stage onwards in *X. tropicalis*

Total RNA from uninjected *X. tropicalis* embryos from stage 8 to 40 was extracted and probed for xtr-miR-F with an antisense LNA probe. RNA sizes (nt) were determined by comparison to radiolabelled decade markers (Ambion).

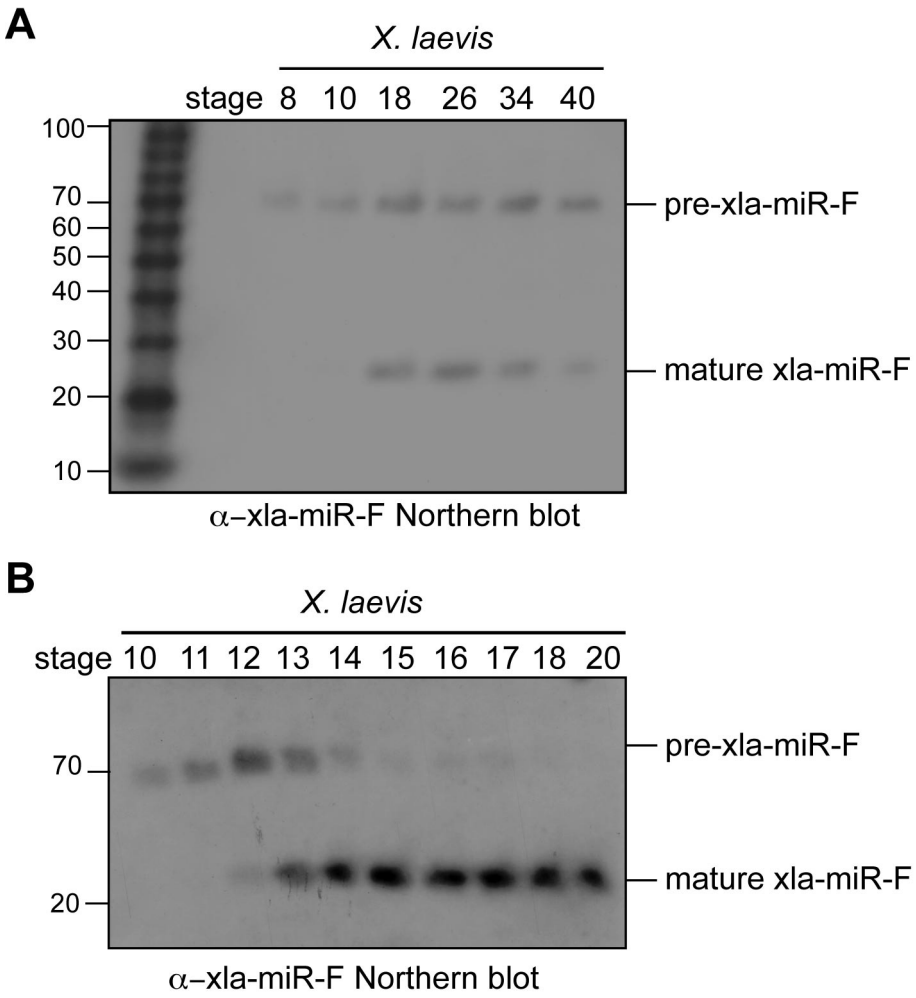


Figure 5.5 miR-F is conserved in *X. laevis* and expressed from the onset of neurulation

A. Northern blot analysis of total RNA from uninjected *X. laevis* embryos at stages 8 to 40 probed for xla-miR-F using an antisense LNA probe. RNA sizes (nt) were determined by comparison to radiolabelled decade markers (Ambion). **B.** Northern blot analysis of total RNA from uninjected *X. laevis* embryos at stages 10 to 20. An antisense DNA oligo was used to probe for xla-miR-F and sizes of small RNAs were determined by comparison to decade markers as in **A.**

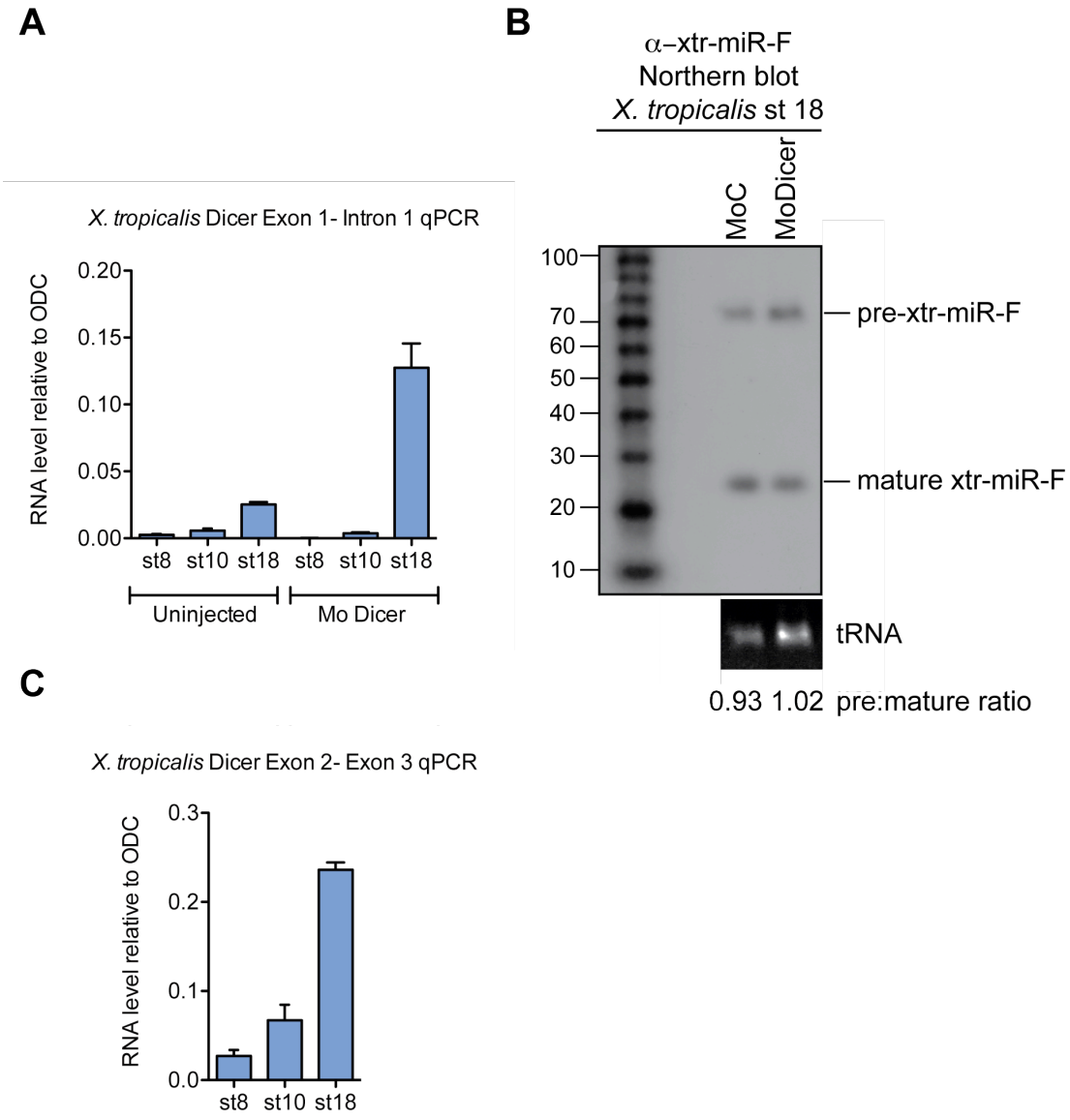


Figure 5.6 xtr-miR-F Dicer-dependency at stage 18 in *X. tropicalis*

A. Total RNA from stage 8, stage 10 and stage 18 *X. tropicalis* embryos either uninjected or injected with 15 ng of a splice-blocking morpholino targeting Dicer (Mo Dicer) was prepared for qPCR. The splice-blocking morpholino is designed to inhibit splicing out of intron 1 from between exons 1 and 2, which will introduce a premature stop codon in Dicer mRNA. Primers complementary to exon 1 and intron 1 of *X. tropicalis* Dicer were used in qPCR amplification. Relative RNA levels to ODC are shown and error bars are standard deviations of triplicate repeats. **B.** Northern blot analysis of total RNA prepared from stage 18 *X. tropicalis* embryos injected with 15 ng of control morpholino (MoC) or Dicer morpholino (Mo Dicer) for xtr-miR-F. Ethidium bromide staining of tRNA serves as a loading control. Pre-miRNA: mature miRNA ratios were quantified in ImageJ. Sizes of small RNAs were determined by comparison to decade markers (Ambion). **C.** Total RNA from uninjected *X. tropicalis* embryos was prepared for qPCR. Primers complementary to exon 2 and exon 3 of *X. tropicalis* Dicer were used in qPCR amplification.

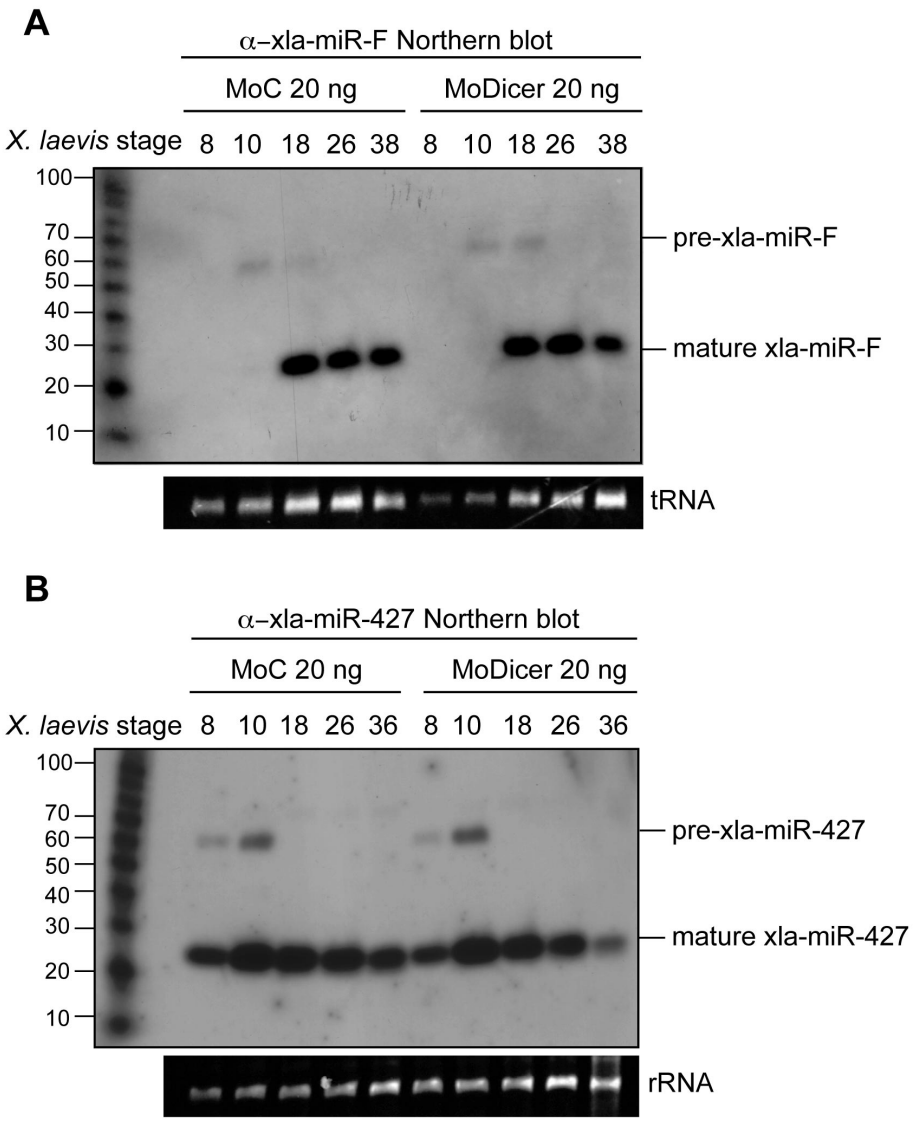


Figure 5.7 Dicer-dependency of xla-miR-F and xla-miR-427 in *X. laevis*

A. Total RNA was prepared from *X. laevis* embryos injected with 20 ng of control morpholino (MoC) or translation-blocking Dicer morpholino (MoDicer) at stages 8 to 38. Northern blot analysis for xla-miR-F using an antisense DNA probe was performed. Ethidium bromide staining of tRNA serves as a loading control. Sizes of small RNAs were determined by comparison to decade markers (Ambion). **B.** Total RNA was prepared from *X. laevis* embryos injected with 20 ng of control morpholino (MoC) or translation-blocking Dicer morpholino (MoDicer) at stages 8 to 36. Northern blot analysis for xla-miR-427 was performed. Ethidium bromide staining of rRNA serves as a loading control.

Figure 5.8 Phenotype of xtr-miR-F knockdown in *X. tropicalis*

X. tropicalis embryos were either uninjected or injected with 15 ng of morpholino against xtr-miR-F (Mo xtr-miR-F) and analysed for phenotypic differences at stage 38. The number of embryos with the phenotype presented out of the total analysed is indicated.

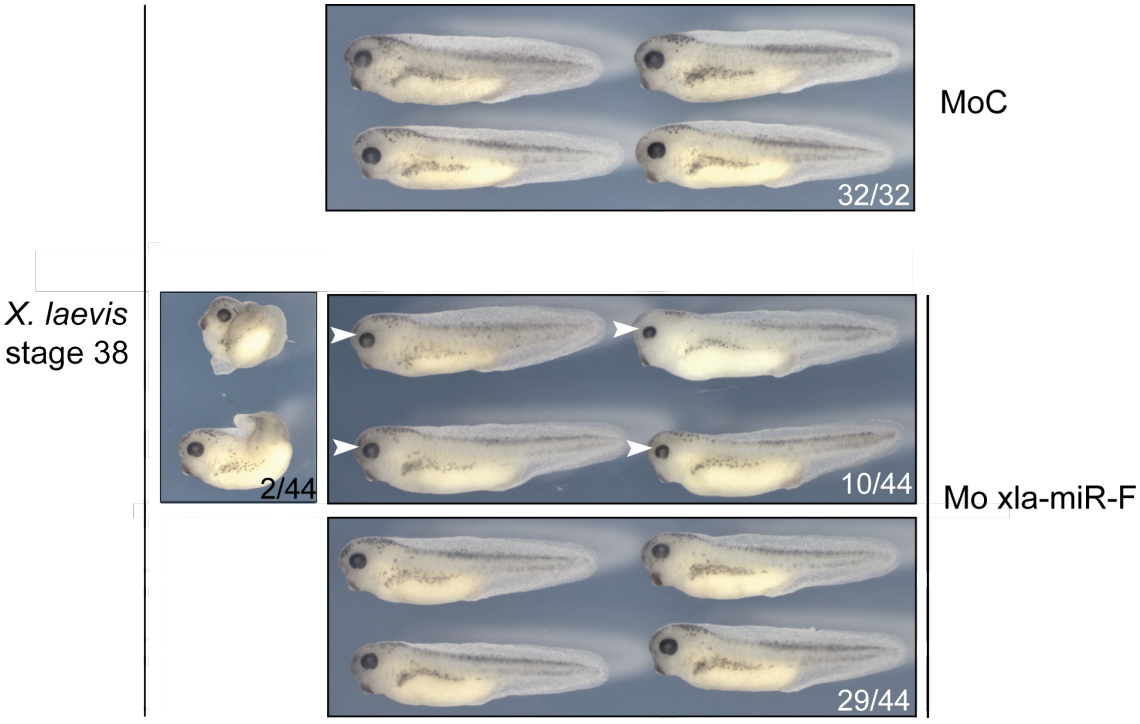


Figure 5.9 Phenotype of xla-miR-F knockdown in *X. laevis*

X. laevis embryos were injected with 15 ng of control morpholino (MoC) or morpholino against xla-miR-F (Mo xla-miR-F) and analysed for phenotypic differences at stage 38. The number of embryos with the phenotype presented out of the total analysed is indicated. White arrowheads indicate reduced eye size in xla-miR-F morphant embryos.

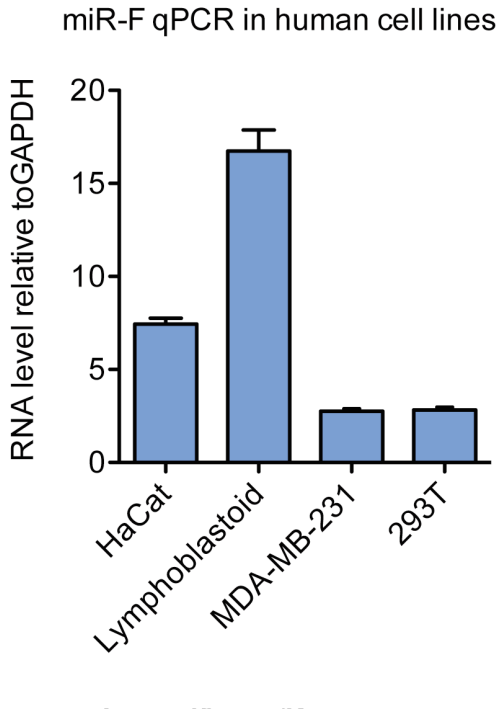


Figure 5.10 miR-F is conserved in human cell lines and down-regulated in a cancer cell line

Total RNA was prepared for miRNA qPCR from the indicated human cell lines. miR-F was amplified and relative RNA levels to GAPDH were calculated. Error bars are standard deviations of triplicate repeats.

5.2.4 Characterisation of unannotated reads

As well as being fed into the novel microRNA prediction pipeline, the unannotated reads that comprise 95 % to 98 % of the *X. tropicalis* early development small RNAome were characterised by a genome-wide annotation of where they align in the genome and their proximity to other small RNAs. For example, unannotated small RNAs could align to a cluster of small RNAs in the genome or be isolated from other small RNAs in the genome. The annotation parameters are defined in the following sections.

5.2.4.1 Definition of Unique, Multi and Intermediate tags

Due to the high multiplicity observed within the small RNA libraries (Table 4.2), unannotated reads were compressed into tags and classified depending on the number of loci in the genome the tag aligned to. All tags aligned to the genome with zero mismatches and were represented by at least 10 identical reads to eliminate noise. Unique tags aligned to a single locus in the genome. Multi tags aligned to ten or more different loci in the genome. Intermediate tags aligned to between two and nine loci in the genome. As shown in Table 5.2, approximately 42 % of tags in all small RNA libraries were classed as unique tags, 25 % were multi tags and 33 % were intermediate tags.

Small RNA library	% Unique tags	% Multi tags	% Intermediate tags
Stage 8	42.9	24.0	33.2
Stage 10	42.1	25.0	33.0
Stage 18	42.7	24.5	32.8
Stage 10 animal	41.0	25.9	33.1
Stage 10 vegetal	43.1	24.5	32.4

Table 5.2 Percentage of tags that are classed as unique, intermediate or multi tags

5.2.4.2 Definition of Low, High and Mixed blocks

The regions of the genome where tags aligned were defined as blocks depending upon the type of tag aligning and how close the aligned tags were to each other (Figure 5.11) (Armisen et al., 2009). Low blocks are a contigous region of the genome containing at least two unique tags that are no more than 200 bases apart, but can contain other tags. High blocks contain at least two multi tags that are no more than 200 bases apart, but

can contain other tags. Mixed blocks contain at least two tags of any type (unique, multi or intermediate) that are less than 200 bases apart. Isolated tags are tags of any type that do not fall into a block. It is worth noting that all high and low blocks will be contained within a mixed block. Mixed blocks are therefore analysed separately to avoid double counting of tags.

Figure 5.12 shows the results of genome-wide small RNA block analysis for tags falling into high, low and mixed blocks or isolated tags. Firstly, very few tags did not fall into any block, as isolated tags were less than 0.08 % of tags in high/low block and mixed block annotations. Isolated tags in the high/low block annotation were exclusively unique tags, as no multi tags fell outside blocks. In other words, very few tags were more than 200 nucleotides from another tag in the genome, and those that were corresponded to small RNAs that only align to one place in the genome (unique tags). Secondly, the percentage of high and low block tags was consistent across the five small RNA libraries. Between 35-37 % of small RNA tags fell into high blocks and between 59-63 % of small RNA tags fell into low blocks.

In the mixed block annotation, intermediate tags are counted in addition to unique and multi tags. The definition of a mixed block is much less stringent than a high or low block, as two tags of any type (multi/unique/intermediate) must be within 200 nucleotides of each other. Mixed blocks therefore overlap some low and high blocks (for example the mixed block in Figure 5.11) and all high and low blocks are contained within a mixed block, therefore mixed block annotation is presented separately in Figure 5.12. Approximately 98 % of tags fall into mixed blocks in all five small RNA libraries.

To fully annotate the unknown 95-98 % of the *X. tropicalis* early development RNAome according to the small RNA block criteria, small RNA tags counted in Figure 5.12 were expanded back into reads and the percentage of reads in each library falling into high, low and mixed blocks was calculated (Figure 5.13). When tags are expanded back into reads, the small RNA block profile is similar in all the libraries and more reads fall into high blocks (56-60 %) than low blocks (35-39 %). This is in contrast to tags, which fell into low blocks more than high blocks (Figure 5.12). This suggests that

multi tags correspond to high numbers of reads as well as aligning to multiple sites in the genome.

When the length distribution of tags falling into low, high, or mixed blocks and isolated tags in each library is analysed (Figure 5.14), similar length distributions for tags are observed in all five small RNA libraries. A 28 nt length peak is seen in all libraries for tags falling into low, high and mixed blocks.

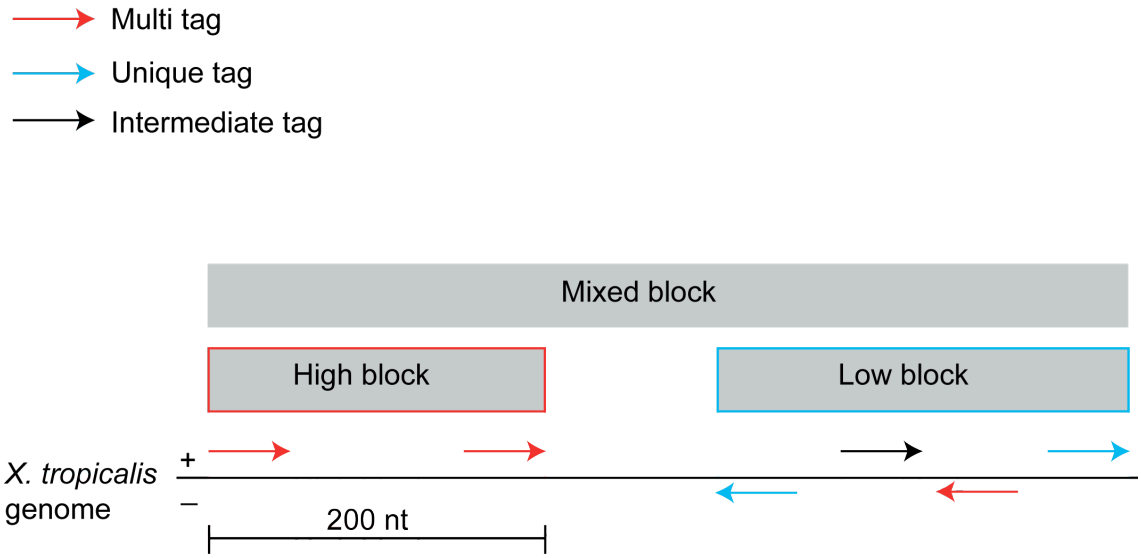


Figure 5.11 Genome-wide small RNA block annotation

Schematic of small RNA block annotation. Multi, unique and intermediate tags align to the *X. tropicalis* genome (plus and minus strand indicated). The minimum criteria for a low or high block are two unique or two multi tags less than 200 nt apart respectively. A minimal high block is shown, containing two multi tags 200 nucleotides apart.

Although the length of the low or high block is determined by the distance between the unique or multi tags, other types of tag may be contained in the blocks. For example, a low block defined by two low tags but also containing a multi tag and an intermediate tag is shown. A mixed block contains two tags of any type that are less than 200 nt apart. For example, a mixed block spans the high and low blocks illustrated as the gap between the high block and the low block is less than 200 nt.

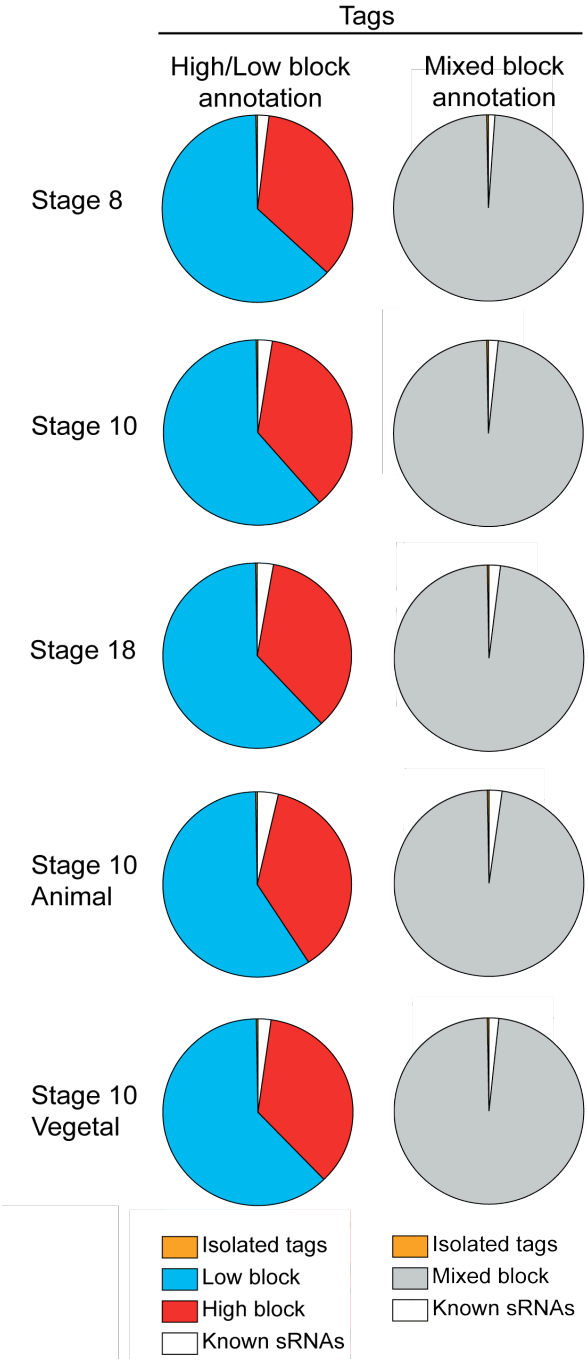


Figure 5.12 Small RNA block annotation of tags

All tags were aligned to the genome and low, high and mixed blocks of tags were defined depending on the type of tag contained within the block (Figure 5.11). Isolated tags did not fall into any block and were less than 0.08 % of tags, hence are not visible on the pie charts. Only multi or unique tags falling into a high or low block, isolated tags and the known small RNA tags identified in Chapter 4 are counted in the high/low block annotation. For the mixed block annotation, all tags that align to the genome and fall into a mixed block, are isolated, or correspond to known small RNA tags are counted. The percentage of annotated tags in each small RNA library is shown.

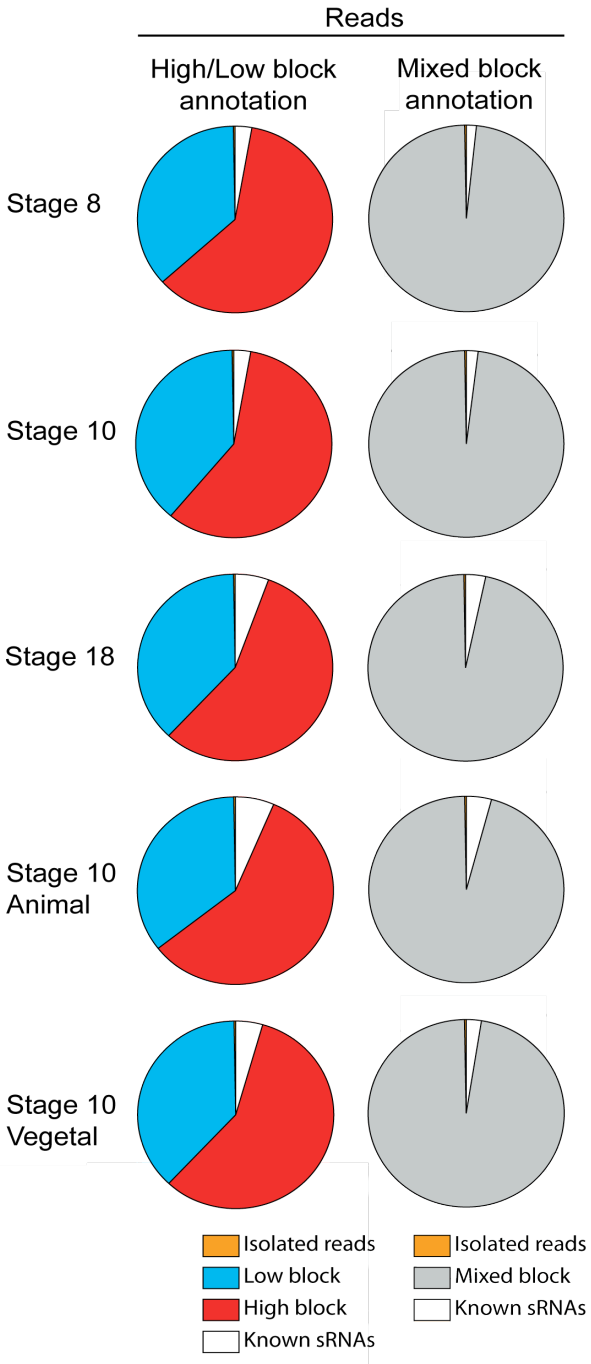


Figure 5.13 Small RNA block annotation of reads

Small RNA block annotation was performed as in Figure 5.12 with all tags expanded back into read numbers. The percentage of annotated reads in each small RNA library is shown. Again, isolated reads are not visible as they were less than 0.09 % of reads.

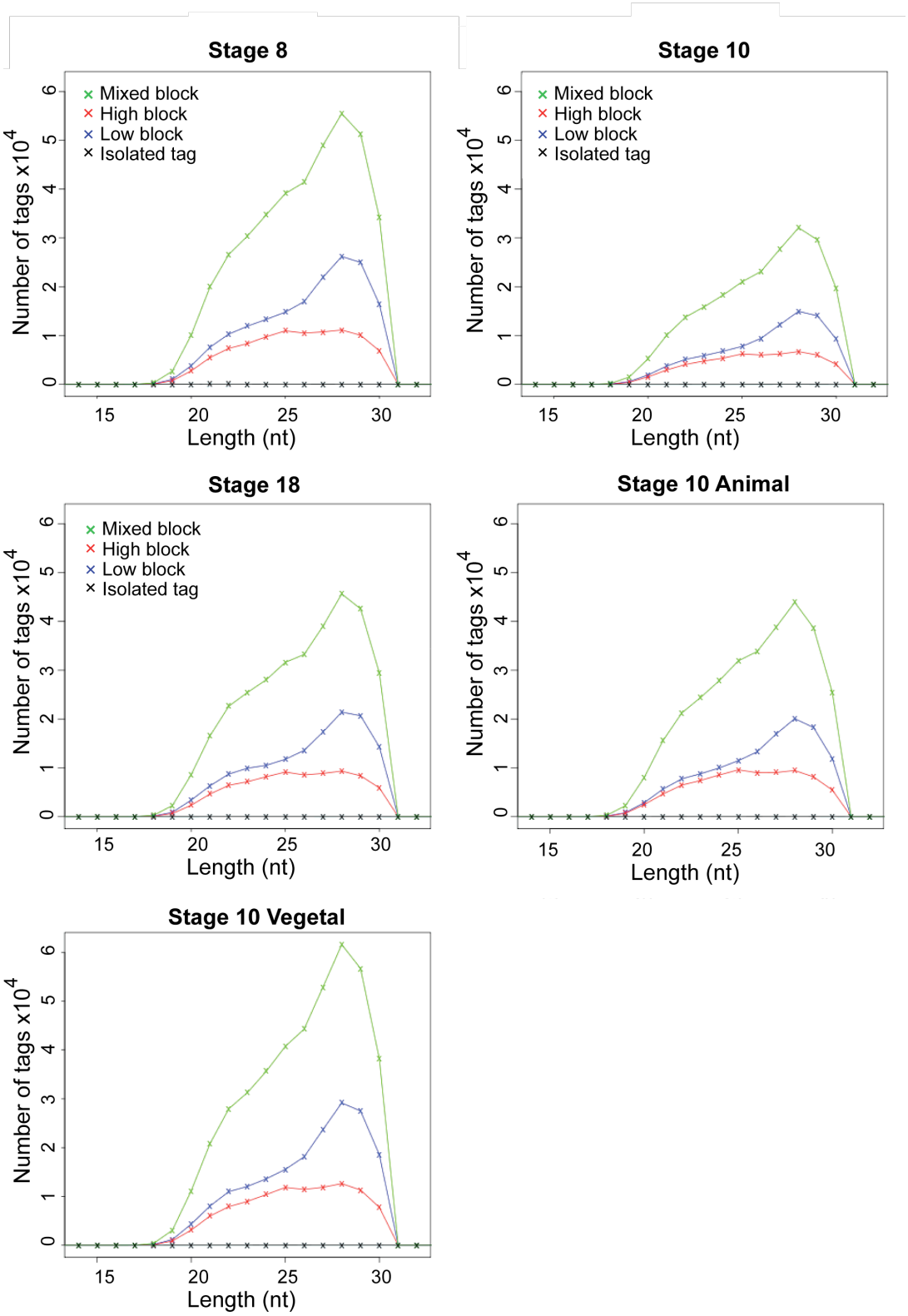


Figure 5.14 Length distributions of tags in small RNA blocks

The length and abundance of small RNA tags that are annotated as falling into a mixed block, high block, low block or as isolated tags is plotted for each small RNA library.

5.2.5 Multi-tags align to introns

The genome-wide small RNA block annotation of the previously unknown small RNA reads characterises the abundance, length, multiplicity and clustering of the unknown reads. Next, I investigated the extent to which the previously unannotated small RNAs align to genes to explore a possible role of the unknown small RNAs in gene regulation.

Firstly, the unannotated tags were analysed for whether or not they aligned to the coding region of a gene as defined by the RefSeq annotation database (Pruitt et al., 2009) (Figure 5.15). Approximately 46 % of tags aligned to a RefSeq coding region (CDS) that includes 5' and 3' UTRs, exons and introns, whereas the remaining 54 % of tags aligned to intergenic regions in all five small RNA libraries. When the tags aligning to coding regions were classified as multi, unique or intermediate according to the definitions in Section 5.2.4.1, it was clear that intermediate and multi tags align to genes more often than unique tags. Intermediate and multi tags aligning to a CDS were about 26 % and 17 % of the unannotated tags, whereas unique tags aligning to a CDS were approximately 3 % of unannotated tags across the small RNA libraries.

Next, tags were analysed for their alignment to introns and exons (Table 5.3). Strikingly, the unannotated tags were found to align much more often to introns than exons, and multi and intermediate tags showed a strong bias towards intron alignment. Multi tags were enriched in introns by at least 40-fold (animal library) compared to exons, and intermediate tags were enriched in introns by at least 20-fold (vegetal library) compared to exons. Both sense and antisense alignments to introns and exons were permitted. 51 % of multi tags aligned to introns in the sense orientation and 49 % in the antisense orientation. 54 % of intermediate tags aligned in the sense orientation and 46 % in the antisense orientation. This result is exciting because it suggests that the unannotated small RNAs cluster in introns, and may therefore be a new class of regulatory small RNAs. Since the small RNAs align in both the sense and antisense orientations in similar proportions and are intermediate and multi tags, this is a strong argument that they are not simply cellular mRNA degradation products, as these would by definition be unique tags that align in the sense orientation.

Figure 5.16 shows previously unannotated small RNAs clustering in the introns of *trim37*, a predicted E3 ubiquitin ligase in *X. tropicalis* (Bowes et al., 2010) in the Integrative Genomics Viewer (IGV) (Robinson et al., 2011). Two large clusters and several minor small RNA clusters align to introns in both the sense and antisense orientations. Strikingly, the clusters are entirely reproducible in all five small RNA libraries and none of the small RNAs align to exons. Further examples of *X. tropicalis* genes with intronic small RNA clusters are illustrated in Figure 5.17. Multiple introns are hit by a small RNA cluster in many of the genes analysed. Figure 5.18 shows an example of an intergenic small RNA cluster upstream of the *top2A* gene, which also contains intronic small RNA clusters. Intergenic small RNA clusters upstream of genic regions were observed for multiple genes, including *igfbp4* and *stat1*.

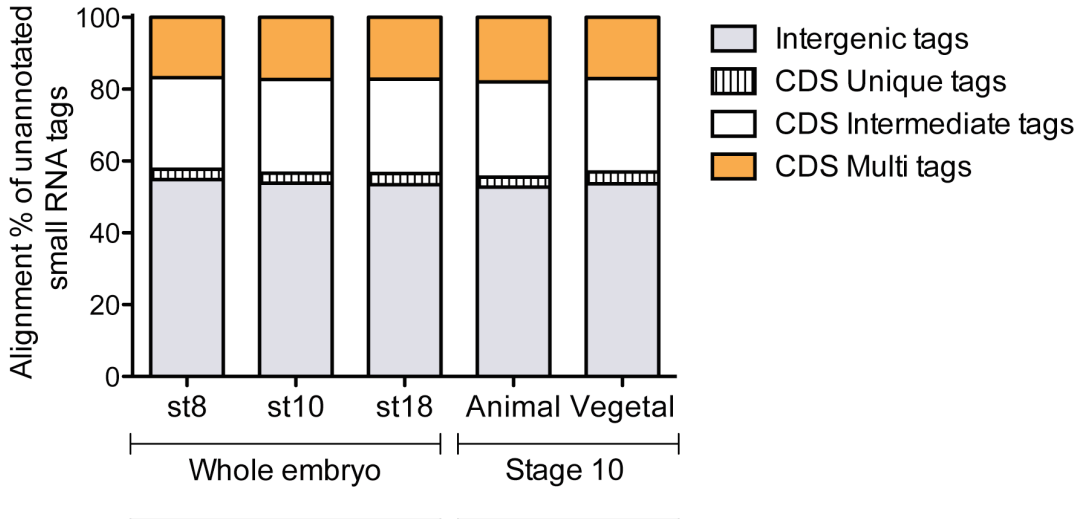


Figure 5.15 Unannotated tags align to coding regions of genes and intergenic regions

All unannotated tags that aligned to the genome with zero mismatches were analysed for alignment to a RefSeq coding region sequence (CDS), which includes 5' and 3'UTRs, exons and introns. Tags that aligned to a CDS were annotated as unique, intermediate or multi tags according to the definitions in Section 5.2.4.2. Tags that did not align to a CDS were classed as intergenic. Tags were allowed to align to a CDS in both the sense and antisense orientations.

For example, if a tag aligned to 10 loci in the genome (a multi tag), and one of those loci was in a CDS, then the tag would be counted as a CDS multi tag. If a tag only aligned to loci in intergenic regions, then it was counted as an intergenic tag.

Small RNA library	% Multi exon tags	% Multi intron tags	% Unique exon tags	% Unique intron tags	% Intermediate exon tags	% Intermediate intron tags
Stage 8	0.73	32.2	0.67	3.66	1.54	45.3
Stage 10	0.82	33.2	0.77	3.60	1.74	46.4
Stage 18	0.84	33.1	0.86	3.91	1.85	46.5
Animal	0.86	34.6	0.78	3.43	1.80	47.4
Vegetal	0.80	32.7	1.27	3.68	2.25	45.4

Table 5.3 Percentage of unannotated tags that align to exons and introns

The number of multi, unique and intermediate tag hits aligning to exons and introns is displayed as a percentage of the total number of unannotated tags. Both sense and antisense tag alignments to exons and introns are permitted. Note that these categories are not mutually exclusive as one tag can align to an exon and an intron at different loci and contribute to the percentage for both categories. Exons are defined by RefSeq, and introns are defined as sequences between two flanking RefSeq exons.

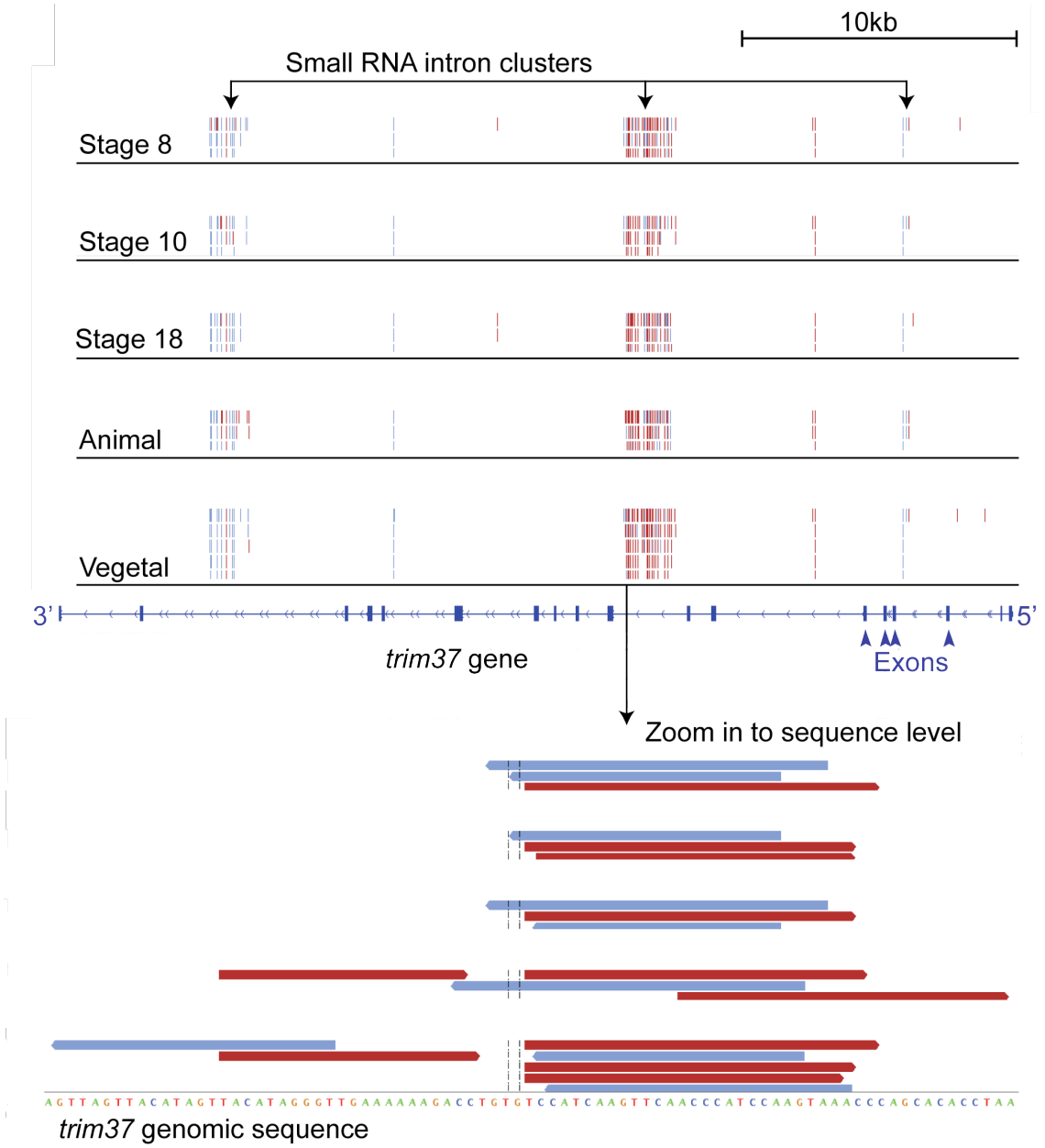


Figure 5.16 Small RNA intron clusters in *trim37*

Unannotated small RNA tags in the stage 8, stage 10, stage 18, stage 10 animal and stage 10 vegetal libraries are displayed aligning to the *X. tropicalis* genome in the Integrative Genomics Viewer (IGV). The gene *trim37* is used as an example of a gene where the unannotated small RNAs cluster in introns. Tags are coloured red for sense alignments and blue for antisense alignments. The entire *trim37* gene is displayed with exons as blue boxes (highlighted with arrowheads) and introns as blue lines. The orientation of the gene is indicated by the arrows on the introns. A 10 kb scale bar is shown. A zoomed in picture of tags in part of a small RNA intron cluster is shown at a sequence level view.

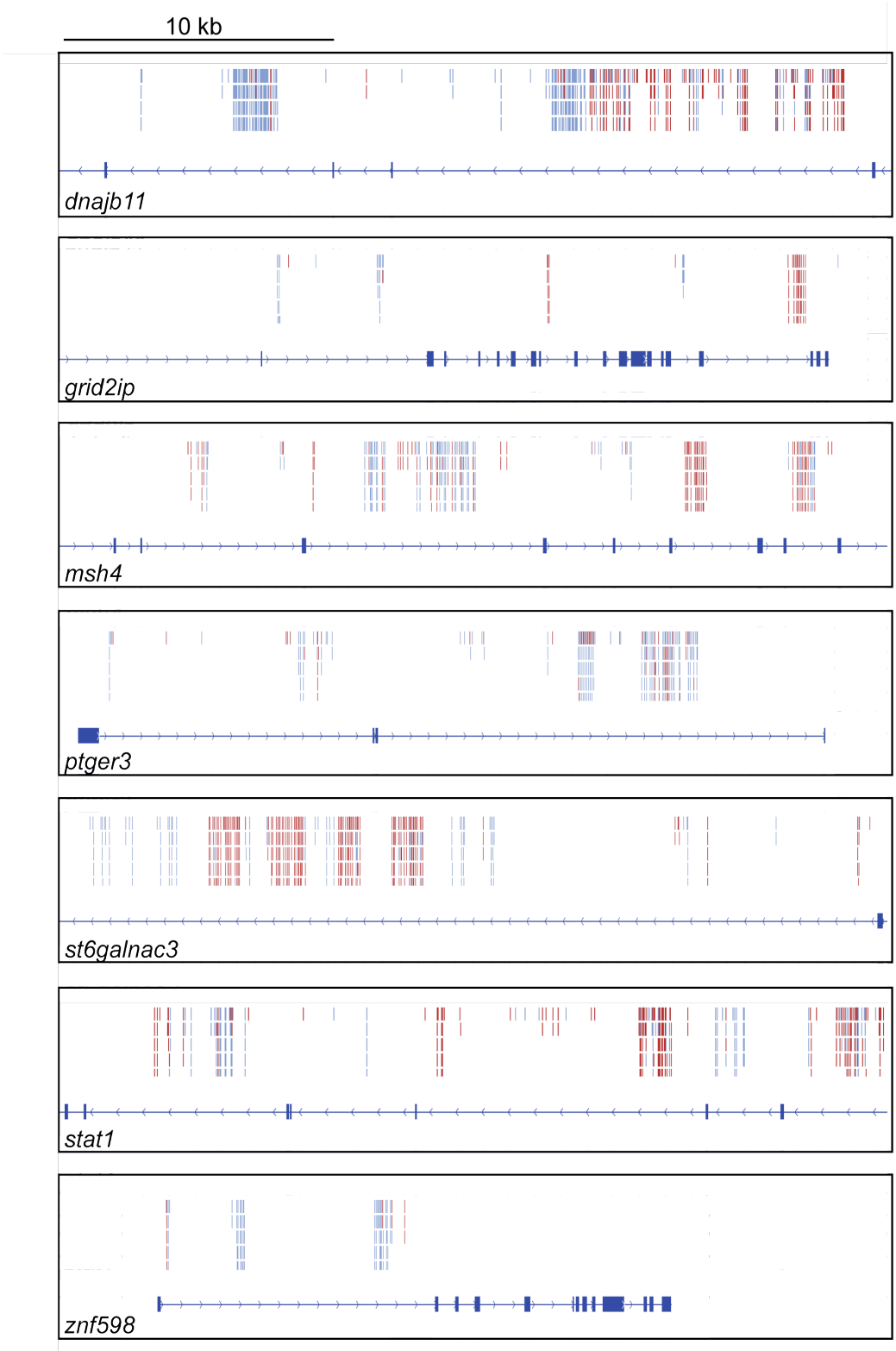


Figure 5.17 Examples of small RNA intron clusters

Small RNA intron clusters in the selected genes are displayed in the IGV browser as in Figure 5.16. A partial or complete gene sequence is displayed for clarity. Only the stage 10 vegetal library track for each gene is shown, but the small RNA intron clusters were conserved in all five small RNA libraries.

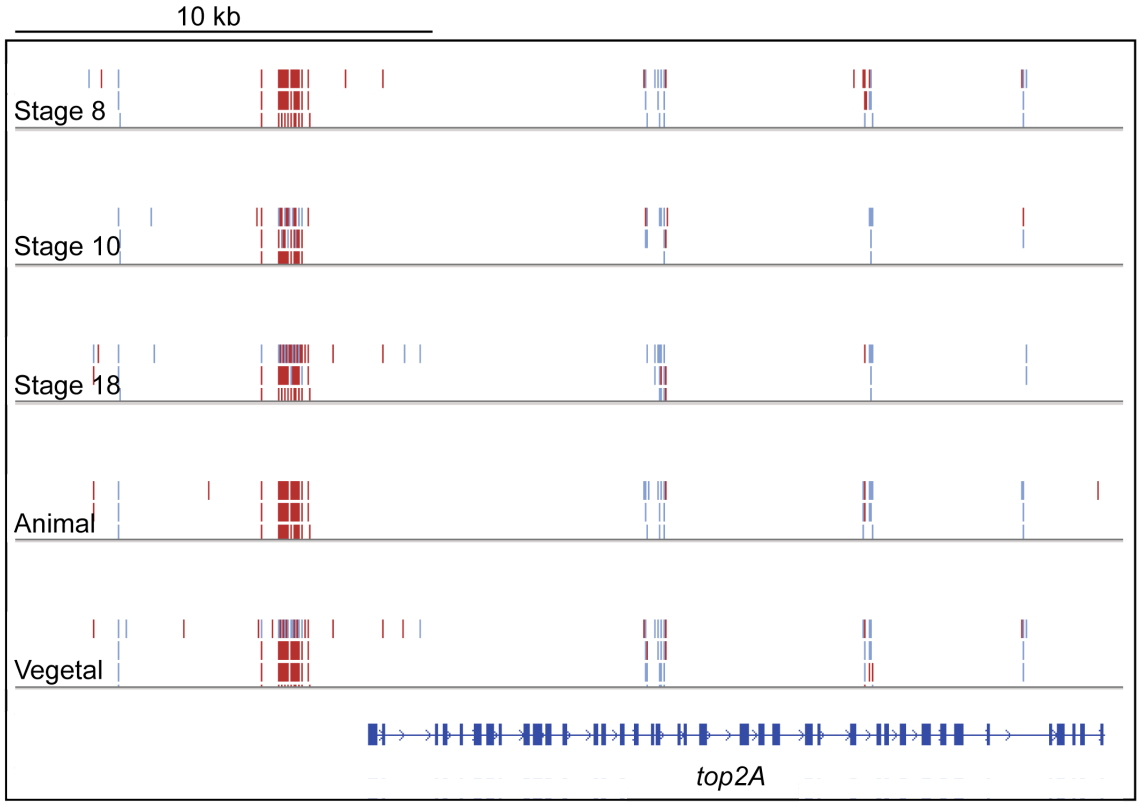


Figure 5.18 Intergenic and intronic small RNA clusters in *top2A*
Small RNA clusters in *top2A* are displayed in the IGV browser as in Figure 5.16. The small RNAs cluster in intergenic regions upstream of the top2A gene and in introns. The small RNA clusters are conserved in the *X. tropicalis* stage 8, 10, 18 and stage 10 animal and vegetal small RNA libraries. A 10 kb scale bar is shown.

5.2.6 Correlation of small RNA intron clusters with gene expression by mRNA sequencing

Small RNAs such as siRNAs, microRNAs and piRNAs negatively regulate gene expression post-transcription (see section 1.4). Moreover, small RNAs can silence gene expression by remodelling chromatin to a transcriptionally inactive state in *Arabidopsis*, yeast, *Drosophila* and mammalian systems (Gonzalez et al., 2008; Pal-Bhadra et al., 2004; Swiezewski et al., 2007; Verdel et al., 2004). In order to investigate whether there was a correlation between the presence of an intronic small RNA cluster in a gene and its expression level, Illumina sequencing of mRNA libraries from the same stages in *X. tropicalis* development as the small RNA libraries was performed.

mRNA sequencing gives a readout of transcript levels which is superior in sensitivity for differential gene expression in comparison to microarrays (Marioni et al., 2008). As described in section 2.7, I extracted total RNA from *X. tropicalis* embryos obtained by *in vitro* fertilisation at stages 8, 10 and 18 and mRNA libraries were prepared by the Advanced Sequencing Facility at the Cancer Research UK London Research Institute. The quality of the mRNA sequencing was confirmed by alignment of the mRNA-seq reads to the *X. tropicalis genome* (Table 5.4) and assessing read coverage of exons of developmental genes that are known to be expressed at stages 8, 10 and 18.

The quality of the alignment of mRNA-seq reads back to the genome was not as high as the small RNA-seq alignment: between 25-28 % of mRNA-seq reads aligned to the genome with two mismatches, in contrast to 49-59 % of small RNA-seq reads with zero mismatches. However, this does not take into account mRNA-seq reads that align over an exon-exon boundary. Including these spliced mRNA reads increased the mRNA-seq alignment to between 35-39 %. The mRNA seq reads aligned to exons as expected, and the expression of *ODC*, *Gsc* and *MyoD* by mRNA-seq matched their known expression profiles (Figure 5.19). *ODC* is expressed at the same level as stages 8, 10 and 18 in *X. tropicalis*, (Paris et al., 1988) and this is observed by the high numbers of mRNA-seq reads aligning in the sense orientation to exons of *ODC* in all the mRNA-seq libraries. *Goosecoid* is highly expressed at stage 10 and declines in expression post-stage 16 (Cho

et al., 1991), and this is mirrored in the mRNA-seq read alignments. *MyoD1* is expressed from stage 14 onwards in *Xenopus (Hopwood et al., 1989)*, and this is reflected by no mRNA seq reads aligning in the stage 8 and 10 mRNA libraries and many reads aligning in the stage 18 library. The overall percentage of genes undergoing at least a 5-fold or 10-fold change in gene expression by mRNA-seq between stages 8, 10 and 18 was analysed to gain an impression of the proportion of genes undergoing developmental regulation in the early *X. tropicalis* embryo (Table 5.5). 31 % of genes increased or decreased in transcript levels by at least 5-fold, and 19.6 % by at least 10-fold, indicating that 31 % of genes undergo substantial developmental regulation.

mRNA library	Number of raw mRNA-seq reads	% of raw reads that align to the genome (2 mismatches)	% of raw reads aligning to the genome including spliced reads
Stage 8	26,595,676	25.4	39.3
Stage 10	25,522,643	25.4	35.3
Stage 18	26,716,900	28.1	38.2

Table 5.4 Overview of mRNA sequencing quality

mRNA sequencing reads from stage 8, stage 10 and stage 18 *X. tropicalis* whole embryos were aligned to version 4.1 of the *X. tropicalis* genome. The number of raw reads obtained in each library and the percentage of raw reads aligning to the genome is shown. Note these are unspliced alignments that do not include mRNA-seq reads over an exon-exon boundary.

mRNA expression	% of genes
At least 5-fold change	31.0
At least 10-fold change	19.6

Table 5.5 Developmentally-regulated gene expression measured by mRNA sequencing

mRNA sequencing reads from stage 8, stage 10 and stage 18 mRNA libraries were normalised to allow comparison of gene expression across the libraries and analysed for fold changes in mRNA-seq read numbers. The percentage of genes undergoing at least a 5-fold or a 10-fold change (increase or decrease) in gene expression between stages 8, 10 and 18 is shown. To account for genes where one stage had zero mRNA-seq reads, zero reads was set at 1 read for calculation of fold changes.

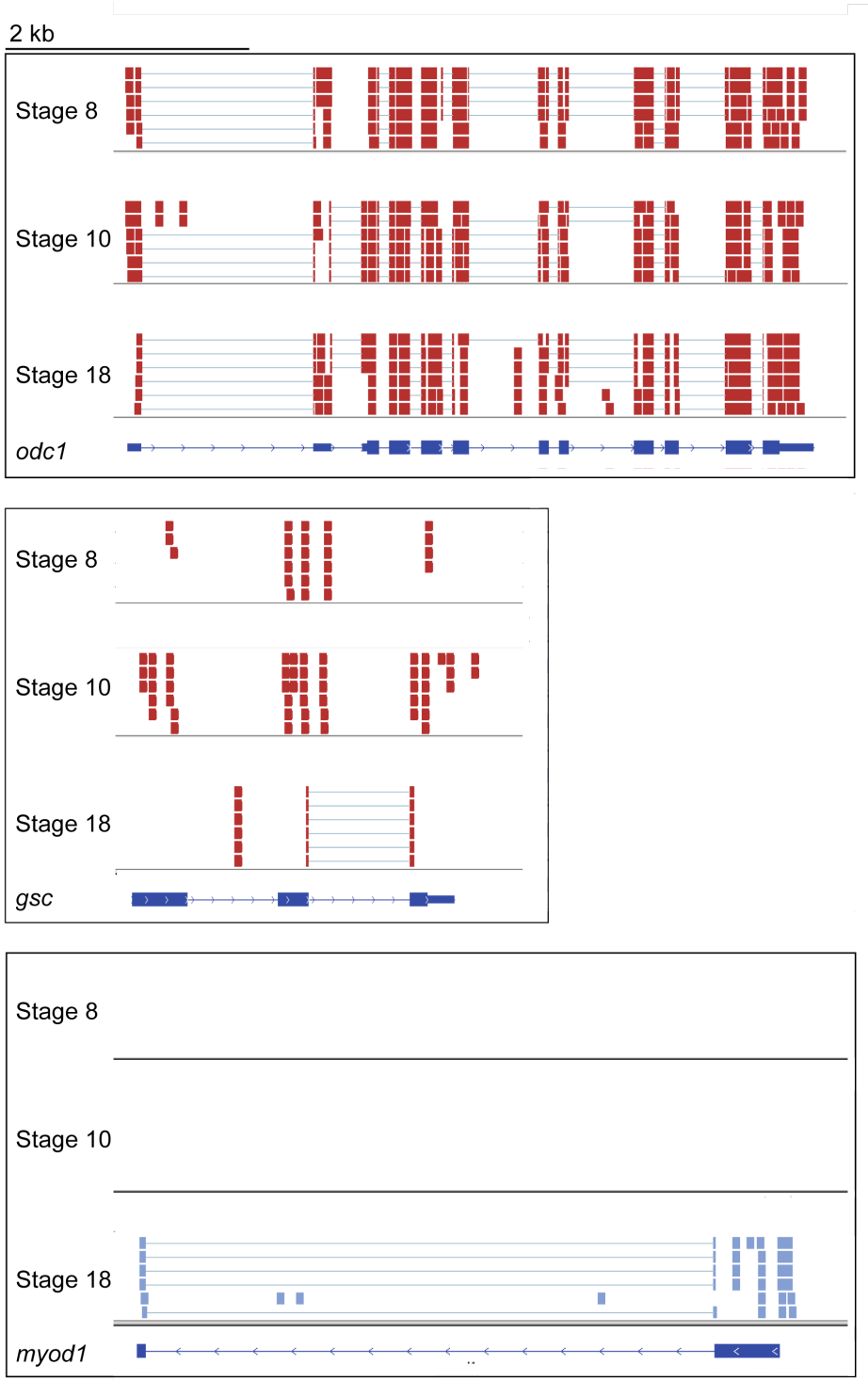


Figure 5.19 mRNA-seq read alignments to *odc1*, *goosecoid* and *myod1* are exonic and match their known expression profiles

mRNA seq read alignment to the exons of *odc1*, *goosecoid (Gsc)* and *myod1* in the stage 8, stage 10 and stage 18 mRNA libraries. IGV browser annotation is the same as Figure 5.16. A 2 kb scale bar for all genes is shown. Spliced mRNA-seq reads are indicated by a blue line connecting reads that span exon-exon boundaries.

Since multi tag small RNAs were the most enriched in introns compared to unique and intermediate tags, for subsequent analysis I focused on multi-tag small RNA intron clusters. In total, 6991 *X. tropicalis* genes were hit by a multi-tag in an intron at stage 8, 6860 genes were hit at stage 10 and 7165 genes were hit at stage 18 (Table 5.6). Multi tags were allowed to align in either the sense or antisense orientation. When the overlap of all genes hit by intronic small RNAs was compared between stages 8, 10 and 18, strikingly, 88 % of the genes were common to all three libraries (Figure 5.20).

To determine whether a gene was 'on' or 'off' by mRNA-seq, I defined 'on' as at least 50 % of exons contained at least one mRNA seq read alignment in the sense orientation and 'off' if the criteria for 'on' were not met (i.e. less than 50 % of exons were hit by an mRNA-seq read). As described in Table 5.6, approximately 63 % of all genes were 'on' at stages 8, 10 and 18. When genes containing an intronic multi tag hit were analysed, a very similar percentage of genes were 'on' and 'off' compared to all genes (~ 63 % and 37 % respectively), suggesting that gene expression is not affected by the presence of an intronic small RNA hit and that many intronic small RNAs are located in genes that are actively transcribed. Intronic small RNAs aligned to 'on' and 'off' genes in the sense and antisense orientations in approximately equal proportions (Table 5.7).

Genes with intronic small RNA hits were then sorted by the mean number of small RNA hits per intron for further analysis (Table 5.8). Although there appears to be no global correlation between the presence of intronic small RNAs and gene expression (Table 5.6), when the genes hit by intronic small RNAs are sorted by highest intronic small RNA density, 70 % of the top 20 genes are 'off' and 66 % of the top 50 genes are 'off', indicating that genes with the most intronic small RNAs may be silenced. This is compared to 37 % of all genes containing an intronic small RNA hit being 'off'. The gene categories hit by small RNAs in Table 5.8 include histones, chromatin-associated proteins, RING-domain containing proteins, membrane proteins (transporters and receptors) and metabolism and signalling proteins. This suggests that intronic small RNAs may silence the expression of these genes. 7 of the top 50 genes changed from 'off' to 'on' or vice versa between stages 8 and 18, indicating that the expression of

most genes with high intronic small RNA density is not developmentally regulated between late-blastula and neurula stage.

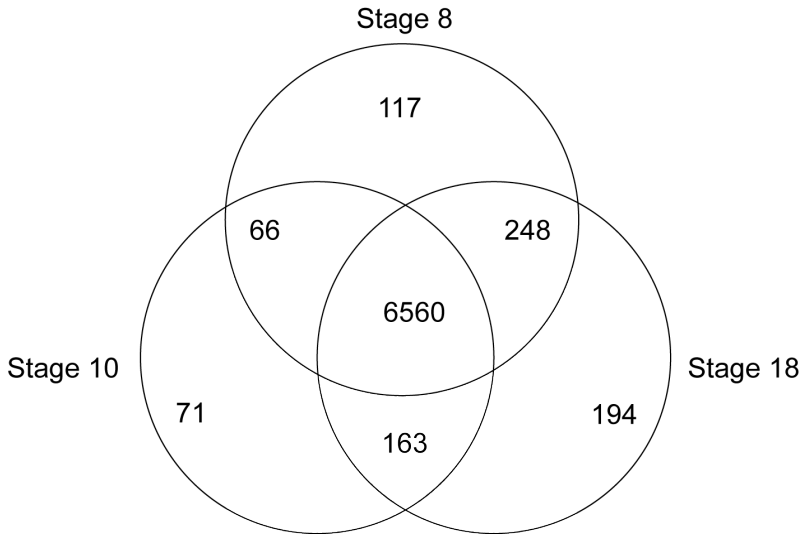


Figure 5.20 Overlap of genes hit by intronic small RNAs

Venn diagram of genes hit by intronic small RNAs at stage 8, stage 10 and stage 18 in *X. tropicalis*.

mRNA library	% all genes ON	% all genes OFF	Number of genes hit by intronic small RNAs	% intronic hit genes ON	% intronic hit genes OFF
Stage 8	62.8	37.2	6991	63.3	36.7
Stage 10	63.8	36.2	6860	63.9	36.1
Stage 18	62.5	37.5	7165	61.9	38.1

Table 5.6 Correlation of small RNA intron clusters and gene expression

Genes were classed as 'on' if at least 50 % of exons contained at least one mRNA seq read alignment in the sense orientation and 'off' if the criteria for 'on' were not met. The 'on' or 'off' status of all genes is compared to genes with at least one multi tag intronic small RNA (intronic hit genes).

mRNA library	% of ON genes with sense intronic small RNAs	% of ON genes with antisense intronic small RNAs	% of OFF genes with sense intronic small RNAs	% of ON genes with antisense intronic small RNAs
Stage 8	50.6	49.4	52.2	49.4
Stage 10	51.0	49.0	52.2	49.0
Stage 18	50.7	49.3	52.1	49.3

Table 5.7 Percentage of sense and antisense intronic small RNAs in ON and OFF genes

Genes that were classified as 'on' or 'off' were analysed for whether the intronic small RNAs reads were aligning in the sense or antisense orientation. Note that in this table, 'on' is defined as at least one mRNA seq read aligning to an exon and 'off' is defined as zero mRNA seq reads aligning to exons.

Gene	Mean small RNA hits per intron			Gene 'on' or 'off' by mRNA-seq		
	st 8	st 10	st 18	st 8	st 10	st 18
histone cluster 2 H2ab	440,429	322,159	311,823	ON	ON	ON
zinc binding alcohol dehydrogenase domain	292,776	297,161	233,488	OFF	OFF	OFF
bromodomain and PHD finger containing 3	252,899	228,102	201,725	OFF	OFF	OFF
SH2B adaptor protein 3	249,408	209,543	202,728	ON	OFF	OFF
myosin 15A_isoform1	226,892	184,136	174,395	OFF	OFF	OFF
interferon-induced protein with tetratricopeptide repeats 5	154,045	201,024	131,011	OFF	OFF	OFF
transmembrane protein 11	146,025	92,939	101,487	ON	ON	ON
solute carrier family 47	142,675	154,358	107,829	OFF	OFF	OFF
myosin 15A_isoform5	131,859	137,462	102,097	OFF	OFF	OFF
c16orf57	129,686	158,498	119,128	OFF	OFF	OFF
copper metabolism (Murr1) domain containing 1	108,629	89,929	82,036	ON	ON	ON
butyrophilin subfamily 2 member A1	99,908	74,720	76,182	ON	ON	OFF
C1q and tumor necrosis factor related protein 4	91,375	99,362	76,198	OFF	OFF	OFF
solute carrier family 39 (zinc transporter), member 3	90,311	91,575	75,337	OFF	OFF	OFF
phosphoribosylformylglycinamidine synthase	85,971	62,323	66,045	OFF	OFF	OFF
ring finger protein 5	85,625	71,398	63,881	ON	ON	ON
hypothetical protein LOC100145503	82,294	94,414	68,400	OFF	OFF	OFF
Cdk5 and Abl enzyme substrate 2	81,805	94,102	67,726	OFF	OFF	OFF
KIAA0895 protein	80,850	78,774	63,053	OFF	OFF	OFF
histone cluster 1, H4a	78,096	44,922	53,412	OFF	OFF	OFF
LY6/PLAUR domain containing 6	73,118	108,350	72,031	OFF	OFF	ON

opioid binding protein/cell adhesion molecule-like	68,940	82,189	71,997	OFF	OFF	OFF
brain abundant membrane attached signal protein	68,210	83,176	49,073	OFF	OFF	ON
chromodomain helicase DNA binding protein 9	65,253	81,976	65,657	OFF	OFF	OFF
sirtuin 4	64,148	88,764	57,847	OFF	OFF	OFF
DNA-directed DNA polymerase kappa	63,440	74,628	46,308	ON	ON	OFF
hypothetical protein LOC100170434	58,576	35,705	38,469	OFF	OFF	OFF
yippee-like 2	58,213	65,725	46,950	OFF	OFF	OFF
small G protein signaling modulator 1	57,450	68,699	42,758	OFF	OFF	OFF
microphthalmia-associated transcription factor	54,893	38,810	40,095	OFF	OFF	OFF
ornithine decarboxylase antizyme 2	53,943	44,475	40,039	ON	ON	ON
novel C3HC4 type (RING finger) and B-box zinc finger protein with SPRY domain	53,482	42,491	36,430	OFF	OFF	OFF
U3 small nucleolar ribonucleoprotein homolog	50,650	40,562	37,754	ON	ON	ON
guanylate binding protein 2	50,551	41,749	37,775	OFF	ON	OFF
chromobox homolog 6	49,334	44,218	37,897	ON	ON	OFF
t-SNARE domain containing 1	48,279	54,001	36,018	OFF	OFF	OFF
interferon-induced protein with tetratricopeptide repeats 5	43,356	42,465	40,060	OFF	OFF	OFF
cell adhesion molecule 2 precursor	42,396	27,970	31,093	OFF	OFF	OFF
solute carrier family 7, member 6 opposite strand	41,838	26,768	29,529	ON	ON	ON
arylsulfatase D	39,807	51,362	36,997	OFF	OFF	OFF
deleted in malignant brain tumors 1_isoform1	38,899	54,236	35,062	ON	OFF	OFF
deleted in malignant brain tumors 1 _isoform2	38,899	54,236	35,062	OFF	OFF	OFF
cytochrome P450 family 2	37,226	20,440	24,742	OFF	OFF	OFF

ring finger protein 126	36,894	30,931	27,714	ON	OFF	OFF
interferon gamma receptor 1	36,732	27,298	28,467	ON	ON	ON
c17orf103	36,458	23,194	25,321	ON	ON	OFF
synapse differentiation-inducing 1	36,093	37,787	26,533	OFF	OFF	OFF
der-like domain containing 2	34,492	28,879	25,725	ON	ON	ON
beta-transducin repeat-containing	34,192	28,703	25,620	ON	ON	ON
crystallin, lambda 1	32,370	26,973	24,439	ON	ON	OFF

Table 5.8 Gene expression status of the top 50 genes hit by intronic small RNAs

X. tropicalis genes containing intronic small RNA hits were sorted by the mean number of small RNA hits per intron and 50 genes with the highest intronic small RNA density are displayed. The mean number of small RNA intron hits at stage 8, 10 and 18 are listed, with the corresponding expression status of the gene at stages 8, 10 and 18 as determined by mRNA-seq transcriptome profiling. ‘On’ and ‘off’ are defined as in Table 5.6.

5.3 Discussion

5.3.1 Summary of results

- The 95 % of small RNA reads that were previously unannotated were fed into a novel microRNA prediction pipeline, with a 50 % experimental validation rate to yield three novel microRNAs, xtr-miR-A, xtr-miR-C and xtr-miR-F.
- xtr-miR-F is expressed from neurula stage to tadpole stage and is conserved in *X. laevis*, where it is expressed from stage 12 onwards.
- Morpholino inhibition of xtr-miR-F causes defects in posterior and axis development, whereas in *X. laevis*, an additional effect on eye development and fewer melanocytes are observed when xla-miR-F is inhibited.
- Dicer mRNA expression increases between stages 10 and 18, similar to global microRNA expression.
- *Xenopus* miR-F is conserved in four human cell lines, and miR-F levels are lower in the breast cancer cell line, MDA-MB-231 compared to untransformed cell lines.
- The unannotated reads were subject to annotation according to their multiplicity in the genome and were classified as multi, unique and intermediate tags.
- Genome-wide clustering of unannotated small RNAs into high and low blocks revealed that more unannotated small RNA reads fell into high blocks, whereas more tags fell into low blocks.
- The majority of unannotated small RNAs that aligned to genes aligned to introns, not exons.
- Clusters of intronic small RNAs were observed in many genes.
- mRNA-seq transcriptional profiling of the stage 8, 10 and 18 *X. tropicalis* embryo allowed a comparison between the presence of small RNA intron clusters and gene expression levels.

- There is no global correlation between gene expression level and the presence of an intronic small RNA.
- Genes with the highest density of small RNA intronic hits are mainly 'off' (66 % of the top 50 genes with intronic hits compared to 37 % of all genes).
- Genes with high intronic small RNA densities are chromatin-associated proteins, RING domain proteins and membrane proteins (receptors and transporters).

The aim of the work presented in this chapter was to characterise and investigate a function for the 95-98 % of the *X. tropicalis* early development small RNAome that could not be annotated by comparison to databases of known small RNAs. I found three novel microRNAs within the unannotated reads, which validates the prediction of microRNAs from small RNA-seq reads based on hairpin-folding criteria. Importantly, I showed that one novel microRNA, xtr-miR-F was conserved in *Xenopus laevis* and in all human cell lines tested. Human miR-F levels are lower in the breast cancer cell line MDA-MB-231 compared to the more normal lymphoblastoid cells and keratinocytes (HaCats), which suggests that it may be worth investigating if miR-F is downregulated in cancer. Global down-regulation of microRNAs is thought to be a feature of cancer (Kumar et al., 2007; Lu et al., 2005) and therefore studying the effect of loss and gain of function of miR-F in human cell lines would potentially reveal whether miR-F plays a role in cancer development, for example as a tumour suppressor. Previous work has shown that microRNAs can act as oncogenes (e.g. the miR-17-92 cluster) (Mendell, 2008), tumour suppressors (e.g. let-7) (Johnson et al., 2005) and initiate metastasis in cancer (miR-10b)(Ma et al., 2007). Moreover, inhibition of the microRNA-processing enzyme Dicer by miR-103/107 has been shown to promote EMT in mammalian systems (Martello et al., 2010b).

X. tropicalis Dicer morphant embryos showed a stabilisation of pre-miR-F and a reduction in mature miR-F at stage 18 (an increase of 10.2 % in the pre-miRNA: mature miRNA ratio, Figure 5.6), suggesting that miR-F is partially Dicer-dependent in *X. tropicalis*. miR-F and xla-miR-427 showed some Dicer-dependency in later stages of *X.*

laevis development (tadpole and tailbud stage respectively), as a reduction in mature miRNA levels was observed at these stages when Dicer was inhibited with a morpholino. This suggests that miR-F and xla-miR-427 are stable, which agrees with previous work showing that mature microRNAs have a half-life of days (Baccarini et al., 2011). miR-F is conserved in humans where there is only one Dicer gene (DICER1(Flicek et al., 2011)), therefore siRNA knockdown of DICER1 in mammalian cells and Northern blot analysis for miR-F would reveal whether miR-F is Dicer-dependent and eliminate the possibility that there are multiple Dicer genes in *Xenopus*.

However, not all microRNAs require Dicer for their processing; miR-451 has been reported as an Argonaute-2-dependent microRNA in mouse and zebrafish (Cheloufi et al., 2010b; Cifuentes et al., 2010). The stability of microRNAs and the existence of more Dicer-independent microRNAs may explain the late phenotypes of Dicer knockdown (reduced eye and neural crest development) shown in Figure 3.21. I found that Dicer mRNA levels in *X. tropicalis* vastly increase between stages 10 and 18, reminiscent of global microRNA expansion in diversity and abundance between stages 10 and 18. This suggests that Dicer is developmentally regulated to facilitate increased microRNA processing as development progresses.

As well as uncovering novel microRNAs contained within the unannotated reads, a genome-wide annotation of the unknown reads revealed that most of the reads cluster together in contiguous regions of the genome that are hit by multi tags (high blocks). When the reads are compressed into tags, low blocks are favoured, which suggests that multi tags are high copy number as well as aligning to multiple loci in the genome. The most abundant unknown reads had a tag length of 28 nt. When the unannotated reads were aligned to genes, they showed a strong bias towards intronic alignment in both the sense and antisense orientations. Small RNA intron clusters are present in many genes, and small RNAs often hit multiple introns in a gene. Determining the extent to which the intronic small RNAs align to repetitive sequences and the level of duplication of intronic small RNA binding sites in the genome will be important for investigating the potential function of the intronic small RNAs. Many repetitive elements in the genome such as transposon integration sites direct the formation of regulatory small RNAs such

as piRNAs (Brennecke et al., 2007), and therefore a bias towards repetitive sequence elements would not exclude a functional role for intronic small RNAs.

Previous studies addressing unknown reads obtained in small RNA-seq have also revealed novel small RNAs with a non-random clustering in the genome. For example, 'tiny' RNAs of sizes around 18 nt were found to associate with transcriptional start sites when small RNA-seq data from mammalian cells, chicken embryo and *Drosophila* was analysed (Taft et al., 2009). 'Tiny' RNAs generally correlated with active transcription and the activatory H3K4 mark. 'Tiny' RNAs were suggested to arise from RNA polymerase stalling or backtracking on the promoter, as tiny RNAs originate from the same strand as the promoter. Another study revealed the existence of 19 nt small RNAs in Hela cells that were derived from Dicer-dependent cleavage of tRNA (Cole et al., 2009). Examination of small RNA-seq data from humans, monkeys and rats led to the discovery of 'miRtrons' in mammals (Berezikov et al., 2007). miRtrons are microRNAs that are generated by splicing excision of introns from mRNA instead of Drosha cleavage in the nucleus (Okamura et al., 2007). Potentially, some of the intronic small RNAs I identified could be miRtrons, although the 20-30 nt size range of intronic small RNAs peaks at 28 nt, in contrast to the 22-23 nt size of microRNAs. In short, a wealth of novel small RNAs with roles in gene regulation have been found by further investigation of unannotated small RNA-seq reads.

In order to determine whether the intronic small RNAs I found by small RNA-seq correlated with active gene expression or silencing, transcriptome profiling by mRNA-sequencing was carried out at the same developmental stages in *X. tropicalis* as the small RNA libraries. Small interfering RNAs (siRNAs) can silence gene expression post-transcription in most organisms (notable exceptions are *Xenopus* and zebrafish) and was initially demonstrated in animals in *C. elegans* (Fire et al., 1998). In *C. elegans*, an RNA-dependent RNA polymerase (Ego-1) synthesises antisense 22 nt siRNAs that bind to the target mRNA to trigger post-transcriptional gene silencing (Maniar and Fire, 2011). In contrast to the small RNAs I sequenced, which were a range of sizes from 20-30 nt and aligned in both the sense and antisense orientation to introns, the *C. elegans* siRNAs aligned in the antisense orientation to exons.

An RNA-dependent RNA polymerase is also required for siRNA amplification and siRNA-mediated post-transcriptional gene silencing in plants (but not in vertebrates)(Baulcombe, 2004). Whereas shorter plant siRNAs (21- 22 nt) correlate with target mRNA degradation, long plant siRNAs (24-25 nt) can also mediate gene silencing by directing heterochromatin formation via histone H3K9 methylation and DNA methylation (Hamilton et al., 2002; Zilberman et al., 2003). Small RNA-seq data in Arabidopsis showed that 'heterochromatic' plant siRNAs align to repetitive sequences (e.g. transposons) and are absent from known protein-coding genes (Kasschau et al., 2007). This contrasts with the intronic location of the small RNA clusters I sequenced, although how repetitive the intronic small RNA loci are has not yet been analysed.

Intriguingly, previous *Arabidopsis* small RNA-seq studies have interpreted clusters of siRNAs as possible sites of biogenesis (Zhang et al., 2007). siRNA clusters that align to one strand of the genome are thought to originate from single-stranded mRNA folded into hairpins and processed by Dicer-like enzymes. siRNA clusters that align to both strands of the genome are suggested to originate from dsRNA synthesised by an RNA-dependent RNA polymerase or from overlapping sense and antisense transcripts (overlapping transcription) (Zhang et al., 2007). Furthermore, antisense transcription of the *Arabidopsis Flowering Locus C* 3'UTR (a protein repressor of flowering) produces two siRNAs of length 24 nt and 30 nt that match one strand at the 3' end of the gene and mediate chromatin silencing via H3K9 methylation (Swiezewski et al., 2007). When the genomic region corresponding to the siRNAs is mutated, the *Flowering locus C* gene is de-repressed and flowering is delayed (Swiezewski et al., 2007).

Since the intronic small RNAs I sequenced align to the genome in the sense and antisense orientations in approximately equal proportions (Table 5.7), and are located in a coding region, it is plausible that the intronic small RNAs originate from overlapping transcription and subsequent cleavage. As overlapping transcription can produce siRNAs of different lengths in *Arabidopsis*, this would fit with the intronic small RNA length distribution of 21-30 nt, (peaking at 28 nt) I observe (Figure 5.14). A further example of overlapping transcription contributing to small RNA-mediated epigenetic silencing is miRNA-directed chromatin remodelling of the INK4/ARF locus in

mammalian cells, which requires an antisense transcript overlapping the promoter (Gonzalez et al., 2008). The long, antisense non-coding RNAs ANRIL and Xist are also synthesised by transcription in the opposite direction to their target genes and are known to mediate gene repression of their complementary loci by recruitment of the PRC2 complex and remodelling of chromatin to a transcriptionally silent conformation via H3K9 methylation (Kotake et al., 2011) (Chow and Heard, 2009).

Other examples of small RNAs negatively regulating gene expression at the level of chromatin include siRNA-directed chromatin remodelling in yeast (Verdel et al., 2004). Heterochromatin formation in yeast is directed by the RNAi-Induced Transcriptional Silencing (RITS) complex that contains siRNA bound to Ago2 and a chromodomain chromatin-binding protein (Chp1) and is mediated by methylation of histone H3K9, which recruits the yeast homologue of heterochromatin protein 1(HP1). piRNA silencing of transposons is also thought to have an epigenetic component, as mutating piRNA biogenesis components relieves silencing of genes normally found in heterochromatin (Pal-Bhadra et al., 2004).

With these observations in mind, intronic small RNAs in *X. tropicalis* genes may negatively regulate gene expression, at either the posttranscriptional level (siRNA-like mechanism) or by transcriptional silencing via chromatin remodelling. When the expression of all *X. tropicalis* genes was compared to those with intronic small RNA hits, there was no difference in global gene expression. However, genes with the highest number of small RNA hits per intron were mainly 'off' (70 % of the top 20) and the genes were in non-random categories such as chromatin-related proteins and membrane proteins (transporters or receptors), suggesting that intronic small RNAs may silence these genes. To investigate epigenetic silencing of genes with high numbers of intronic small RNA hits, histone modifications on the genes could be assessed by analysing an *X. tropicalis* genome-wide Chip-Seq dataset where the activatory and repressive chromatin marks (trimethylated lysine 4 on histone H3, and H3K27 respectively) and RNA polymerase II binding were correlated with gene expression by mRNA-seq (Akkers et al., 2009), or by comparing to H3K9 methylation (another activatory histone modification). This would reveal if genes with intronic small RNAs were associated with distinct activatory or repressive chromatin modifications that correlate with their

level of gene expression. Peaks of RNA polymerase II binding near small RNA clusters may indicate a possible mechanism of small RNA biogenesis. The existence of intergenic small RNA clusters upstream of genic regions suggests that the same analysis of mRNA-seq gene expression correlation with the presence of intergenic small RNA clusters could be performed to investigate whether intergenic small RNAs target promoter or enhancer regions to regulate gene expression. In summary, the intronic small RNA clusters in the previously uncharacterised 95-98 % of the *X. tropicalis* early development RNAome merit further investigation as they may be a new class of novel regulatory small RNAs and will be discussed further in Chapter 6.

Chapter 6. Concluding Discussion

The aims of this thesis were to characterise the role of small RNAs as regulators in early vertebrate development, firstly by investigating the mechanism of spatial regulation of *Xenopus Cripto-1*, a coreceptor for the Nodal signalling pathway, and secondly, by genome-wide sequencing of the small RNAome of the early *X. tropicalis* embryo.

6.1 Summary of Results

- Spatial regulation of *Xenopus Cripto-1* translation in the gastrula-stage embryo is required for control of Nodal signalling timing and magnitude during germ layer specification and is critical for normal embryonic development.
- MicroRNA regulation of Nodal signalling occurs at the level of coreceptor via xla-miR-427 targeting of XCR1.
- Spatial regulation of XCR1 via its 3' UTR is partially Dicer-dependent and is mediated by three sequence elements in the 3' UTR of *XCR1α*: 2 necessary spatial elements and an RNA-stabilising element.
- Genome-wide sequencing of the early *X. tropicalis* embryo revealed that there were between ~370,000 and 740,000 unique small RNAs present (Table 4.2), of which more than 95 % were previously uncharacterised in any species. 192 microRNAs were identified in the early *X. tropicalis* embryo, of which 163 were new to *Xenopus* but had been identified in other species. 64 piRNA sequences were also identified.
- MicroRNA expression increases between late-blastula and late-neurula stages, with microRNAs expanding in both diversity and abundance, which correlates with increased Dicer mRNA expression. MicroRNAs showed a bias towards expression in the animal pole (prospective ectoderm) in gastrula-stage embryos.
- In contrast, stable piRNA expression is observed between blastula and neurula stages with no bias towards animal or vegetal pole localisation in gastrula-stage embryos.
- Characterisation of a novel microRNA, miR-F, uncovered a role in posterior, axial, eye and melanocyte development in *Xenopus*, and conservation in human

cell lines with lower miR-F levels detected in a breast cancer cell line compared with untransformed cell lines.

- Characterisation of the unannotated small RNA reads led to the identification of novel intronic small RNA clusters, which may silence expression of genes with the highest intronic small RNA density as revealed by transcriptome profiling of the early embryo. These genes include chromatin-associated proteins, RING-domain proteins and membrane proteins.

6.2 Implications of the results and future work

6.2.1 Regulation of Cripto-1 and Nodal signalling in development and cancer

An unbiased mapping approach led me to identify 3 critical elements in the XCR1 3' UTR required for spatial regulation. Utilising the list of *X. tropicalis* microRNAs expressed at the same stages of development as XCR1 spatial regulation from my genome-wide small RNA sequencing led to the identification of a critical 5' TATTAG sequence beginning at nucleotide 360 in the 3' UTR that reverts spatial regulation when mutated. However, the microRNA or RNA-binding protein regulators that interact with this site, or the RNA-stabilising and second spatial elements remain uncharacterised. Spatial regulation occurred at the level of translation, showed 3' UTR-dependency and partial Dicer-dependency, indicating a role for microRNA regulation. As discussed in section 3.3.4, potential spatial regulators interacting with the 5' TATTAG sequence include the RNA-binding protein QKI (of which there are two pseudoalleles in *X. laevis*, *qki-a* and *qki-b*) and xtr-miR-187 (mmu-miR-187 was sequenced at stage 10). Moreover, the RNA-binding protein Hur/ELAV has multiple predicted sites in the minimal 3' UTR sequence. QKI and Hur/ELAV can promote RNA stability and QKI can also act as a translational repressor, which could explain repression of XCR1 translation in the vegetal pole.

In order to test whether these proteins or miRNAs spatially regulate XCR1, loss of function studies by morpholino coinjection with GFP-3' UTR reporter mRNAs to assay spatial regulation should be carried out. Determining the expression profiles of these candidate regulators, and their localisation in the embryo at stages 8 and 10 may reveal

whether spatial localisation of an RNA-binding protein or mature microRNA contributes to the mechanism of XCR1 spatial regulation. Interplay between RNA-binding proteins and microRNAs may also be important, as several studies have indicated the proximity of Hur, Pumilio and DND1 sites to microRNA binding sites and the influence of RNA-binding proteins on RNA secondary structure alterations or miRNA binding site occlusion (Mukherjee et al., 2011) (Kedde et al., 2007; Kedde et al., 2010). RNAi in *Xenopus* and microRNA overexpression by injection of double-stranded precursors and Ago mRNA should now be possible due recent work showing that limiting Ago protein prevents RNAi in *Xenopus* (Lund et al., 2011). The effect of Ago protein loss and gain of function on spatial regulation would also be illuminating, as loss of Ago function should rescue spatial regulation if microRNAs are responsible. Ago gain of function should enhance any miRNA-mediated repression. Interestingly, the interaction of Ago with GW182 proteins (Eulalio et al., 2008) suggests another mechanism of miRNA-mediated translational repression of XCR1 in the vegetal pole, as GW182 proteins can inhibit translation initiation via Ago recruitment to the 3' UTR.

I showed that XCR1 in the animal cap was negatively regulated by xla-miR-427, a critical regulator of embryogenesis in zebrafish and *Xenopus* and of cell fate in human embryonic stem cells. MicroRNAs therefore regulate Nodal signalling in early vertebrate development at the level of Nodal coreceptor expression, as well as targeting ligands, receptors and inhibitors in zebrafish and frogs. xla-miR-427 activity did not affect spatial regulation or target the 3' UTR of *XCR1*. However a seed match site in the coding region remains to be tested, which could be done by injection of a xla-miR-427 target protector complementary to the coding region miR-427 site. The human miR-427 homologue, miR-302 only targets the Nodal inhibitor Lefty in human ES cells, thereby upregulating Nodal signalling. Since Nodal is secreted by melanomas and signalling is reactivated in prostate cancer, it would be interesting to profile the expression of miR-302 in melanoma and prostate cancer cells, to explore whether miR-302 function is involved in aberrant Nodal signalling in these cancers. A search for Lefty expression in cancer in the Oncomine database (Rhodes et al., 2007) reveals downregulation of *Lefty* mRNA in B-cell lymphoma and upregulation of *Lefty* mRNA in oesophageal cancer and testicular carcinoma and seminoma.

Cripto-1 is overexpressed in many cancers and is a stem cell marker, consistent with its expression in early development when embryonic cells are pluripotent. Since Cripto-1 is repressed by xla-miR-427 and xla-miR-427 promotes differentiation into mesodermal cell fates, it would be intriguing to analyse whether loss of miR-427/302 promotes Cripto-1 expression in cancer and inhibits differentiation, a feature of tumour cells (Hanahan and Weinberg, 2000). The potential for 3' UTR mediated microRNA regulation of human *Cripto-1* has not yet been fully explored, as I tested whether regulators present in the *Xenopus* embryo could confer spatial regulation of the human 3' UTR (Figure 3.13). MicroRNA regulation of human *Cripto-1* could be investigated in human ES cells (where *Cripto-1* is expressed) using luciferase-3'UTR reporters with mutated microRNA binding sites. A survey of predicted *Cripto-1* microRNA binding sites in TargetScan (www.targetscan.org) reveals three highly conserved 8 nt exact microRNA seed matches in the 3' UTR; hsa-miR-516a-3p, hsa-miR-299-5p and hsa-miR-338-5p.

6.2.2 Small RNA regulation in early vertebrate development

I performed a genome-wide small RNA sequencing study of the early *X. tropicalis* embryo when it became clear from working on the mechanism of spatial regulation of XCR1 that very few microRNAs had been identified in early *Xenopus laevis* development, as most of the known miRNAs had been cloned in adult *X. tropicalis* tissues. I showed that microRNAs increase in abundance and diversity from late-blastula to neurula stage, agreeing with previous studies of global microRNA expression in early zebrafish development. This result is also consistent with the role of microRNAs as regulators of differentiation, as the embryo becomes less pluripotent with the specification of more tissue types between gastrulation and neurulation, and with global microRNA downregulation in cancer leading to inhibition of tumour cell differentiation. Moreover, microRNAs were enriched in the animal pole compared to the vegetal pole at gastrulation, which may contribute to the differentiation of neural tissue in the prospective ectoderm, as neural-specific expression of microRNAs has been widely reported (Walker and Harland, 2008; Wienholds et al., 2005).

MicroRNAs have been sequenced in unicellular organisms such as *C. reinhardtii* (Molnar et al., 2007), suggesting that they existed before complex organisms evolved,

however, a comparative microRNA profiling study reveals that microRNAs are more diverse in complex animals compared to simpler animals such as *Nematostella* (anemone), suggesting that microRNA expansion is linked to evolutionary progression (Grimson et al., 2008). Moreover, the lesser severity of miRNA knockouts in animals such as *C. elegans* compared to the mouse (Smibert and Lai, 2008) suggests that microRNA control of gene expression is essential in vertebrates.

I also showed that 64 piRNA sequences were present in the early *X. tropicalis* embryo and that piRNAs were less abundant and dynamic than microRNAs. The main function of piRNAs is thought to be suppression of transposons in the germline to preserve genomic integrity for subsequent gamete production (Lin and Spradling, 1997). Recently, other studies have sequenced piRNAs in somatic tissues (e.g. mouse pancreas) (Yan et al., 2011), suggesting that piRNAs exist and may have a function outside of germline development. Since piRNAs were not sequenced in adult *X. tropicalis* tissues in a previous study (Armisen et al., 2009), it is possible that piRNA transcription is not maintained into adulthood as the embryo develops. The function of piRNAs in somatic cells could also be to silence non-transposon repeats and maintain heterochromatin repression of repetitive elements in the genome.

In order to take the small RNA sequencing dataset forward to further explore the role of small RNAs in early vertebrate development, the function of the most abundant and spatially localised microRNAs and piRNAs could be studied in more detail. For example, morpholino loss of function or dsRNA overexpression (gain of function) experiments could be performed for microRNAs where the mammalian homologues are not well-characterised, for example, cfa-miR-30e, the second most abundant microRNA I sequenced or hsa-miR-1825, which strikingly accounts for 50 % of miRNAs reads in the vegetal pole at stage 10. Full experimental validation of the piRNA sequences should be performed, firstly to confirm a biochemical interaction with Piwi proteins and the presence of the 3' 2-*O*-methyl modification, a distinguishing feature of piRNAs. The presence of endo-siRNAs in *X. tropicalis* was suggested by the presence of 21- 24 nt small RNA species in the small RNA-seq dataset, but this was not characterised further. A comparison of *X. tropicalis* 21-24 nt small RNA sequences to the recently

identified endo-siRNAs in mouse ES cells could be used to profile endo-siRNA expression in early *X. tropicalis* development (Babiarz et al., 2008).

By examination of the previously unannotated small RNA reads, I identified 3 novel microRNAs and characterised one, miR-F, in detail. This microRNA will be named by sequence similarity on submission to the miRbase microRNA database. The expression profile and phenotype of miR-F depletion in *Xenopus* morphants suggested roles in axial, posterior, eye and melanocyte development. *In situ* hybridisations for molecular markers of neural development (e.g. N-tubulin) and neural crest development (Slug and Snail) may give more insight on the mechanism of miR-F activity in early development. Since miR-F is conserved in mammalian cells, miR-F knockdown or overexpression in cell culture followed by mRNA-seq would identify direct mRNA targets without the complications of multiple tissue interactions in the *Xenopus* embryo. More extensive profiling of miR-F in normal and cancer cell lines should be carried out to determine whether the lower levels of miR-F observed in the breast cancer cell line MDA-MB-231 compared to untransformed cell lines is a more general occurrence, and if so, the question of the role of miR-F in tumourigenesis could be investigated.

I also identified a novel class of intronic small RNAs by examining where the unannotated reads aligned in the *X. tropicalis* genome. Firstly, the genomic sequence that the small RNAs align to could be compared to the Repbase database of eukaryotic transposon, LINE, SINE and LTR and other genomic repeats (Jurka et al., 2005) to assess if some of the small RNA target sequences are derived from transposon integration or other repeats. This could illuminate the function of this novel class of small RNAs, as repeat-associated RNAs can be involved in transposon silencing, or heterochromatic silencing of repetitive regions (see sections 1.5.1 and 1.5.3). Since the small RNAs are 21-28 nt in length, map to introns in the sense and antisense orientation and align to genic regions, it is possible that the small RNAs are derived from overlapping sense and antisense transcription and subsequent cleavage into small RNAs, in a mechanism similar to endo-siRNA biogenesis. This could be tested by RT-PCR amplification of antisense transcripts and analysing the Dicer and Ago2 dependency of the small RNAs. Also, to further investigate the question of where and how the intronic small RNAs are made, RNA polymerase II binding near or overlapping

an intronic small RNA cluster could be analysed by comparing my small RNA-seq data with a genome-wide Chip-seq dataset including RNA polymerase II binding in *X. tropicalis* (Akkers et al., 2009). Moreover, the abundance and expression profiles of intronic small RNAs could be analysed by small RNA-qPCR.

Although there was no global correlation of gene expression with the presence of an intronic small RNA cluster by mRNA-seq, 66 % of the 50 genes with the highest number of intronic small RNA hits were 'off' compared to 37 % of all genes with intronic hits being 'off', suggesting that intronic small RNAs could negatively regulate the genes they align to. This could be via a mechanism involving chromatin-mediated epigenetic silencing. This could be tested by looking for repressive chromatin modifications on genes hit by intronic small RNAs. For example, chromatin modifications could be analysed by comparison to a genome-wide Chip-seq dataset for activatory H3K4 trimethylation and repressive H3K27 trimethylation deposition in *X. tropicalis* (Akkers et al., 2009), or by Chip-Seq for repressive marks such as trimethylated H3K9 or HP-1 binding. This analysis could be extended to include the intergenic small RNA clusters I observed (Figure 5.18) to investigate a possible role in gene regulation. In short, the unanswered questions regarding the novel intronic small RNAs are where are they synthesised from in the genome, what regulates their biogenesis and most importantly, what effect, if any, do they have on gene regulation? The presence of histones, chromatin-associated proteins and RING-finger containing proteins in the genes with the highest density of intronic small RNAs is suggestive of a role in chromatin remodelling, which remains to be explored. Other genes were membrane proteins and 'housekeeping' genes, suggesting that intronic small RNAs target non-random categories of genes. This may indicate a role for intronic small RNAs in regulating timing of replication or spatial control of gene expression in the nucleus (Sexton et al., 2007).

6.2.3 Implications of the genome-wide sequencing approach

The genome-wide small RNA sequencing and mRNA sequencing datasets provide an important resource cataloguing the small RNAome and gene expression profile of early *X. tropicalis* development that will be available to the research community on publication of this work. 95 % of the *X. tropicalis* early development small RNAome

could not be identified by comparison to known small RNAs, suggesting that there are many small RNAs and possible regulatory mechanisms in the early embryo that we do not yet understand. However, in this study, I characterised the unannotated reads by their abundance and clustering in the genome and experimentally validated a novel microRNA prediction pipeline starting from unannotated reads with a success rate of 50 %, i.e., 50 % of the predicted novel miRNAs were experimentally validated. Also, by analysing the genomic alignment of unannotated small RNAs, I noticed that most small RNAs that align to genes are intronic, and by integration of the small RNA-sequencing with genome-wide mRNA sequencing, the expression of the targeted genes was assessed to explore a possible function of intronic small RNAs in gene regulation. This illustrates the power of combining genome-wide sequencing technologies such as small RNA-seq and mRNA-seq, or Chip-seq and mRNA-seq to correlate histone modifications with gene expression to assess if global mechanisms of gene regulation exist. For example, to investigate whether microRNA or siRNA play a global role in regulation of the chromatin landscape as some studies have suggested, genome-wide Chip-sequencing for histone modifications in Dicer and Ago2 mutant zebrafish compared to wild type animals could be compared.

Small RNA-sequencing in this thesis has raised the questions of what roles do piRNAs, novel microRNAs and intronic small RNAs play in early development. Previous work has shown that pursuing unannotated reads that appear to be distributed non-randomly in the genome can lead to the identification of new classes of small RNAs that give insight onto mechanisms of gene regulation. For example ‘tiny’ RNAs are associated with the promoter of actively transcribed genes in human, chicken and *Drosophila*, where they are suggested to be products of RNA polymerase stalling or backtracking (Taft et al., 2009).

The microRNAs I identified in the early *X. tropicalis* embryo can be used as a list of candidate regulators of any developmental process occurring between late-blastula and late-neurula stage, such as dorso-ventral patterning, EMT in germ layer specification and neural crest migration, or neural patterning. For example, the microRNAs present at stages 8 and 10 in *X. tropicalis* were screened for perfect seed matches to the critical regulatory region of the *Xenopus Cripto-1* 3’ UTR which led me to identify the

necessary element for spatial regulation of XCR1, which is essential for the spatial control, timing and magnitude of Nodal signalling. Another example of how the dataset could be developed is screening for novel microRNA regulators of EMT within the newly identified *X. tropicalis* miRNAs to investigate how microRNAs outside the miR-200 family might regulate EMT in development and in cancer. I have obtained a list of new microRNAs with exact seed matches to the 3'UTRs of the key EMT genes *Snail1*, *Snail2*, *Sox9*, *Sox10*, *Twist1* and *Twist2*, all of which are expressed in the neurula stage embryo. To begin with, GFP-3'UTR wild-type and microRNA-binding site mutant reporters could be analysed for miRNA-mediated repression in the *Xenopus* embryo and the role of microRNA regulation of EMT genes could then be explored in mammalian systems. In summary, this thesis has generated a wealth of data that is the starting point for understanding how small RNAs regulate signalling and patterning embryonic tissues in early vertebrate development.

Appendix

Table 6.1 List of miRbase version 17 microRNAs sequenced in the *X. tropicalis* small RNA libraries			
hsa-miR-4753-3p	rno-miR-3557-5p	dme-miR-994-5p	mcmv-miR-m01-3*
ppt-miR1023e-3p	hsa-miR-30a*	gga-miR-140*	osa-miR166e
gma-miR4395	rno-miR-3596c	cte-miR-184b	sme-miR-2a-2*
aly-miR4239	pma-miR-1b	dme-miR-9c-5p	cin-miR-4001d-5p
rno-miR-465*	ssc-miR-320	dme-miR-995-3p	oar-miR-543-3p
cte-miR-31	bmo-miR-1b*	dsi-miR-1012	gma-miR1520p
pma-miR-25b	hsa-miR-378c	dme-miR-317-3p	oan-miR-454*
bfl-miR-34b-5p	rno-miR-455*	oan-miR-7-1*	sko-miR-4818b*
hsa-miR-452*	mmu-miR-322*	dme-miR-313-5p	tgu-miR-146b*
pab-miR947	mmu-miR-1b-5p	pma-miR-30b	mmu-miR-3084
oan-miR-187*	cel-miR-44-3p	hsa-miR-92a-2*	bdi-miR5199
ppy-miR-576-5p	cte-miR-1	rno-miR-322*	hsa-miR-4768-5p
oan-miR-1421y-2*	dme-miR-279-3p	ssc-miR-1306-5p	dps-miR-2572*
ame-miR-3794	dme-miR-312-3p	dme-miR-1011-3p	vvi-miR3627
rno-miR-3553	hsa-miR-103b	mmu-miR-23b*	mmu-miR-92a-2*
dre-miR-140*	oar-miR-379-5p	crm-lin-4	csi-miR3946
bfl-miR-182c	pma-miR-128	mmu-miR-350	gga-miR-1553*
hsa-miR-4443	pma-miR-1c	ppc-miR-63b	hsa-miR-4645-3p
mmu-miR-669a-5p	dme-miR-34-5p	gga-miR-126	oan-miR-146a
cel-miR-2953-3p	dme-miR-31a-5p	hsa-miR-4692	dme-miR-4-3p

tca-miR-29-3p	rno-miR-291b	hsa-miR-4530	tae-miR5086
hsa-miR-3925-5p	dre-miR-202*	mmu-miR-149*	mmu-miR-674*
mmu-miR-3096-5p	xtr-miR-199a*	zma-miR164d*	mmu-miR-106b*
hsa-miR-4451	dme-miR-311-5p	hsa-miR-4454	mmu-miR-28*
tca-miR-3791-3p	dme-miR-14-5p	oan-miR-92c	tca-miR-100-5p
bmo-miR-3259	mmu-miR-1a	bfl-miR-4896	api-miR-263b
cel-miR-4816	dme-miR-10-5p	sme-bantam-c*	dme-miR-8-5p
ppy-miR-920	hsa-miR-3591-3p	cte-miR-317	rno-miR-3554
dps-miR-2567c*	api-miR-184b	hsa-miR-3196	mmu-miR-3963
pma-miR-16	cel-miR-54-5p	cin-miR-4154-3p	oan-miR-15b
rgl-miR5142	mmu-miR-466i-3p	bhv1-miR-B10	aae-miR-124
tca-miR-2944c-5p	hsa-miR-23b*	bfl-miR-71*	hsa-miR-3074-5p
oar-miR-370-5p	bta-miR-26c	aly-miR4236	ssc-miR-151-3p
hsa-miR-3127-5p	rno-miR-3543	dme-miR-999-3p	ppy-miR-199b-5p
bmo-miR-2779	mmu-miR-3095-3p	mmu-miR-3098-3p	mmu-miR-3074-5p
oar-miR-1185-3p	rno-miR-582*	bta-miR-2889	hsa-miR-202*
cin-miR-4034-5p	bta-miR-3604	tgu-miR-460a*	dme-miR-124-3p
bta-miR-2898	rno-miR-3556b	rno-miR-351*	rno-miR-202
pma-miR-4571-3p	bfl-let-7a*	bta-miR-2890	gga-miR-202*
gma-miR156g	tca-miR-7-5p	bfl-miR-4857	ssc-miR-146b
dps-miR-2574a	rno-miR-92b*	osa-miR2866	hsa-miR-4485
sme-miR-13	mmu-miR-92b*	cin-miR-1-3p	cel-miR-239a*
oan-miR-196a*	pma-miR-194	xtr-miR-181a-2*	aly-miR847*
rno-miR-3573-3p	ssc-miR-365-5p	hsv1-miR-H7	rlcv-miR-rL1-34-3p

aly-miR838	dme-miR-7-5p	gga-miR-449b*	hsa-miR-3907
dme-miR-2501-3p	lgi-miR-96b	mmu-miR-5109	ptr-miR-922
bmo-miR-3242	mmu-miR-669b*	rno-miR-3571	hsa-miR-4497
hsa-miR-33a*	crm-miR-87*	bta-miR-2897	hsa-miR-4459
mmu-miR-140*	dps-miR-2574b	pta-miR783	ppt-miR390c
mmu-miR-1963	crm-miR-35e	nvi-miR-281	bmo-miR-2818
bmo-miR-2840	ath-miR5021	zma-miR482	zma-miR171k*
xtr-miR-181a-1*	hsa-miR-3545-5p	pma-miR-1a	dme-miR-975-5p
dps-miR-2555*	rno-miR-3590-5p	hsa-miR-3656	ssc-miR-2476
pma-miR-143	bta-miR-2957	mmu-miR-5119	mmu-miR-3102-3p.2
zma-miR395a*	cbr-miR-71	api-miR-275	hsa-miR-4762-5p
api-miR-29	rno-miR-3586-3p	aae-miR-275	rno-miR-3542
pma-miR-203a	hsa-miR-499a-5p	dme-miR-275-3p	cbr-miR-2226
hsa-miR-4486	mmu-miR-101c	cqu-miR-8	bmo-miR-3404
aly-miR822*	tgu-miR-456	dme-miR-92a-5p	cte-miR-29b
tca-miR-989-3p	crm-miR-48	pma-miR-192	pma-miR-130a
sko-miR-4819*	mmu-miR-365-1*	aae-miR-92a	tca-miR-3872-5p
ppy-miR-16	pma-miR-29a	tca-miR-9b-3p	ppy-miR-1233
pma-miR-152a	rno-miR-3545-5p	mmu-miR-3102-5p.2	rno-miR-3597-5p
hsa-miR-30c-2*	pma-miR-22b	mmu-miR-5107	hsa-miR-3607-3p
mmu-miR-5110	cte-miR-34	rno-miR-194*	tca-miR-92b-3p
hsa-miR-4732-5p	cte-miR-29a	hsa-miR-378f	dre-miR-1788*
bta-miR-2399*	pma-miR-22a	hsa-miR-194*	tgu-miR-200a*

crm-miR-40	rno-miR-17-1-3p	hsa-miR-19a*	mdo-miR-200a*
rno-miR-208	crm-miR-1	oan-miR-205*	oan-miR-19a*
dps-miR-989	oan-miR-92a-1*	mmu-miR-19a*	rno-miR-203*
mmu-miR-30e*	dme-miR-31b-5p	ssc-miR-199a*	mmu-miR-130b*
cte-miR-2695	bfl-miR-4885a	oan-miR-200b*	ssc-miR-133b
bmo-miR-3312	mmu-miR-3057-3p	cel-miR-77-5p	tgu-miR-200b*
bta-miR-2904	crm-miR-75	rno-miR-3596b	tgu-miR-489
pma-miR-203a*	osa-miR5075	hsa-miR-92b*	rno-miR-133a*
hsa-miR-1224-3p	iltv-miR-I1*	mmu-miR-1934*	hsa-miR-3153
rno-miR-221*	tca-miR-263a-5p	dme-miR-310-3p	rno-miR-34a*
pma-let-7d	cte-miR-1987	dme-miR-311-3p	mmu-miR-301a*
dps-miR-2513b	aae-miR-263a	mmu-miR-18a*	dme-miR-219-5p
hsa-miR-221*	rno-miR-146a*	dme-miR-276a-3p	mmu-miR-3102
dme-miR-989-3p	mtr-miR2634	mmu-miR-363-5p	rno-miR-3552
tca-let-7-5p	cel-miR-55*	dme-miR-304-5p	tgu-miR-34c
far-miR1134	oar-miR-487b-5p	aae-miR-14	aae-miR-13
dme-miR-306-5p	ame-miR-989	dme-miR-278-5p	tgu-miR-1329
bfl-miR-125b	mtr-miR2595	dme-miR-8-3p	dps-miR-2517b
tca-miR-3858-3p	cte-miR-10c	bta-miR-3602	ssc-miR-335
bfl-let-7b	ssc-miR-206	aae-miR-31	rno-miR-3556a
oan-miR-221*	mtr-miR2673a	hsa-miR-3120-5p	xtr-miR-9a*
bta-miR-3600	gga-miR-146b	dps-miR-137*	mmu-miR-205*
csi-miR3949	dme-miR-988-3p	tgu-miR-124*	dre-miR-19a*
rno-miR-3596d	hsa-miR-4286	hsa-miR-4252	dme-miR-210-3p

pma-let-7c	bta-miR-146a	aae-miR-276	hsa-miR-130b*
rno-miR-423*	crm-miR-63*	dme-miR-14-3p	rno-miR-3582
tgu-miR-30b-3p	ppc-miR-63a	dme-miR-79-3p	rno-miR-205
bta-miR-193a	cel-miR-54-3p	hsa-miR-18a*	hsa-miR-34a*
mmu-miR-30c-1*	mmu-miR-378b	aae-miR-317	egr-miR-124b
mmu-miR-25*	hsa-miR-378d	cte-miR-7	xtr-miR-9a
cte-miR-71	mmu-miR-5128	cqu-miR-210	cte-miR-92c
mmu-miR-221*	mmu-miR-1306-5p	dme-miR-263a-5p	oan-miR-205
dme-miR-184-3p	crm-miR-80	hsa-miR-877*	pma-miR-200b
dme-miR-1-3p	dme-miR-970-3p	dre-miR-3906	pma-miR-9a*
dme-let-7-5p	crm-miR-249	tca-miR-31-5p	dme-miR-92a-3p
hsa-miR-25*	mmu-miR-17*	oan-miR-1349	
bta-miR-3596	dme-miR-305-3p	ssc-miR-199b	

Reference List

Adamson, E.D., Minchiotti, G., and Salomon, D.S. (2002). Cripto: a tumor growth factor and more. J Cell Physiol *190*, 267-278.

Adkins, H.B., Bianco, C., Schiffer, S.G., Rayhorn, P., Zafari, M., Cheung, A.E., Orozco, O., Olson, D., De Luca, A., Chen, L.L.*, et al.* (2003). Antibody blockade of the Cripto CFC domain suppresses tumor cell growth in vivo. J Clin Invest *112*, 575-587.

Agius, E., Oelgeschlager, M., Wessely, O., Kemp, C., and De Robertis, E.M. (2000). Endodermal Nodal-related signals and mesoderm induction in Xenopus. Development *127*, 1173-1183.

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Francoijs, K.J., Stunnenberg, H.G., and Veenstra, G.J. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev Cell *17*, 425-434.

Al-Souhibani, N., Al-Ahmadi, W., Hesketh, J.E., Blackshear, P.J., and Khabar, K.S.A. (2010). The RNA-binding zinc-finger protein tristetraprolin regulates AU-rich mRNAs involved in breast cancer-related processes. Oncogene *29*, 4205-4215.

Albrecht, M., and Lengauer, T. (2004). Survey on the PABC recognition motif PAM2. Biochem Biophys Res Commun *316*, 129-138.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M.*, et al.* (2003a). A uniform system for microRNA annotation. RNA *9*, 277-279.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and other tiny endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

Anderson, C., Catoe, H., and Werner, R. (2006). MIR-206 regulates connexin43 expression during skeletal muscle development. Nucleic Acids Res *34*, 5863-5871.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

Armes, N.A., and Smith, J.C. (1997). The ALK-2 and ALK-4 activin receptors transduce distinct mesoderm-inducing signals during early Xenopus development but do not co-operate to establish thresholds. Development *124*, 3797-3804.

Armisen, J., Gilchrist, M.J., Wilczynska, A., Standart, N., and Miska, E.A. (2009). Abundant and dynamically expressed miRNAs, piRNAs, and other small RNAs in the vertebrate Xenopus tropicalis. Genome Res *19*, 1766-1775.

Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z.R., Asai, K., Siomi, H., and Siomi, M.C. (2008). Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A *105*, 7964-7969.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Baccarini, A., Chauhan, H., Gardner, T.J., Jayaprakash, A.D., Sachidanandam, R., and Brown, B.D. (2011). Kinetic Analysis Reveals the Fate of a MicroRNA following Target Regulation in Mammalian Cells. Curr Biol *21*, 369-376.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res *33*, 7138-7150.

Barroso-Deljesus, A., Lucena-Aguilar, G., Sanchez, L., Ligero, G., Gutierrez-Aranda, I., and Menendez, P. (2011). The Nodal inhibitor Lefty is negatively modulated by the microRNA miR-302 in human embryonic stem cells. FASEB J *25*, 1497-1508.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*, 356-363.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev *20*, 1885-1898.

Bennett, J.T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., and Schier, A.F. (2007). Nodal signaling activates differentiation genes during zebrafish gastrulation. Dev Biol *304*, 525-540.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nat Genet *35*, 215-217.

Bianco, C., Strizzi, L., Mancino, M., Rehman, A., Hamada, S., Watanabe, K., De Luca, A., Jones, B., Balogh, G., Russo, J*., et al.* (2006). Identification of cripto-1 as a novel serologic marker for breast and colon cancer. Clin Cancer Res *12*, 5158-5164.

Bianco, C., Strizzi, L., Normanno, N., Khan, N., and Salomon, D.S. (2005). Cripto-1: an oncofetal gene with many faces. Curr Top Dev Biol *67*, 85-133.

Biedermann, B., Hotz, H.R., and Ciosk, R. (2010). The Quaking family of RNA-binding proteins: coordinators of the cell cycle and differentiation. Cell Cycle *9*, 1929-1933.

Biedermann, B., Wright, J., Senften, M., Kalchhauser, I., Sarathy, G., Lee, M.H., and Ciosk, R. (2009). Translational repression of cyclin E prevents precocious mitosis and embryonic gene activation during C. elegans meiosis. Dev Cell *17*, 355-364.

Blanchet, M.H., Le Good, J.A., Mesnard, D., Oorschot, V., Baflast, S., Minchiotti, G., Klumperman, J., and Constam, D.B. (2008a). Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation. EMBO J *27*, 2580-2591.

Blanchet, M.H., Le Good, J.A., Oorschot, V., Baflast, S., Minchiotti, G., Klumperman, J., and Constam, D.B. (2008b). Cripto localizes Nodal at the limiting membrane of early endosomes. Sci Signal *1*, ra13.

Bonev, B., Pisco, A., and Papalopulu, N. (2011). MicroRNA-9 reveals regional diversity of neural progenitors along the anterior-posterior axis. Dev Cell *20*, 19-32.

Bou Kheir, T., Futoma-Kazmierczak, E., Jacobsen, A., Krogh, A., Bardram, L., Hother, C., Gronbaek, K., Federspiel, B., Lund, A.H., and Friis-Hansen, L. (2011). miR-449 inhibits cell proliferation and is down-regulated in gastric cancer. Mol Cancer *10*, 29.

Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A., and Fusenig, N.E. (1988). Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J Cell Biol *106*, 761-771.

Bouwmeester, T., Kim, S., Sasai, Y., Lu, B., and De Robertis, E.M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. Nature *382*, 595-601.

Bowes, J.B., Snyder, K.A., Segerdell, E., Jarabek, C.J., Azam, K., Zorn, A.M., and Vize, P.D. (2010). Xenbase: gene expression and improved integration. Nucleic Acids Res *38*, D607-612.

Bracken, C.P., Gregory, P.A., Kolesnikoff, N., Bert, A.G., Wang, J., Shannon, M.F., and Goodall, G.J. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer Res *68*, 7846-7854.

Brandt, R., Normanno, N., Gullick, W.J., Lin, J.H., Harkins, R., Schneider, D., Jones, B.W., Ciardiello, F., Persico, M.G., Armenante, F.*, et al.* (1994). Identification and biological characterization of an epidermal growth factor-related protein: cripto-1. J Biol Chem *269*, 17320-17328.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brivanlou, A.H., Gage, F.H., Jaenisch, R., Jessell, T., Melton, D., and Rossant, J. (2003). Stem cells. Setting standards for human embryonic stem cells. Science *300*, 913-916.

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep *9*, 582-589.

Cailleau, R., Young, R., Olive, M., and Reeves, W.J., Jr. (1974). Breast tumor cell lines from pleural effusions. J Natl Cancer Inst *53*, 661-674.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Chaffer, C.L., and Weinberg, R.A. (2011). A perspective on cancer cell metastasis. Science *331*, 1559-1564.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010a). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584-589.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010b). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584-589.

Chen, C., and Shen, M.M. (2004). Two modes by which Lefty proteins inhibit Nodal signaling. Curr Biol *14*, 618-624.

Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci *20*, 465-470.

Chen, J.F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., and Wang, D.Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet *38*, 228-233.

Chen, X., Li, Q., Wang, J., Guo, X., Jiang, X., Ren, Z., Weng, C., Sun, G., Wang, X., Liu, Y.*, et al.* (2009). Identification and characterization of novel amphioxus microRNAs by Solexa sequencing. Genome Biol *10*, R78.

Chen, Y.I., Moore, R.E., Ge, H.Y., Young, M.K., Lee, T.D., and Stevens, S.W. (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res *35*, 3928-3944.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E., and Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. Nature *447*, 823-828.

Cheng, A.M., Thisse, B., Thisse, C., and Wright, C.V. (2000). The lefty-related factor Xatv acts as a feedback inhibitor of Nodal signaling in mesoderm induction and L-R axis development in xenopus. Development *127*, 1049-1061.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res *31*, 3497-3500.

Cho, K.W., Blumberg, B., Steinbeisser, H., and De Robertis, E.M. (1991). Molecular nature of Spemann's organizer: the role of the Xenopus homeobox gene goosecoid. Cell *67*, 1111-1120.

Choi, W.-Y., Giraldez, A.J., and Schier, A.F. (2007). Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. Science *318*, 271-274.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol *21*, 359-366.

Christen, B., and Slack, J.M. (1999). Spatial response to fibroblast growth factor signalling in Xenopus embryos. Development *126*, 119-125.

Chung, W.J., Okamura, K., Martin, R., and Lai, E.C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr Biol *18*, 795-802.

Ciardiello, F., Kim, N., Saeki, T., Dono, R., Persico, M.G., Plowman, G.D., Garrigues, J., Radke, S., Todaro, G.J., and Salomon, D.S. (1991). Differential expression of epidermal growth factor-related proteins in human colorectal tumors. Proc Natl Acad Sci U S A *88*, 7792-7796.

Ciccodicola, A., Dono, R., Obici, S., Simeone, A., Zollo, M., and Persico, M.G. (1989). Molecular characterization of a gene of the 'EGF family' expressed in undifferentiated human NTERA2 teratocarcinoma cells. EMBO J *8*, 1987-1991.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D*., et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA *15*, 2147-2160.

Collier, B., Gorgoni, B., Loveridge, C., Cooke, H.J., and Gray, N.K. (2005). The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. EMBO J *24*, 2656-2666.

Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., and van Roy, F. (2001). The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell *7*, 1267-1278.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2010). RBPDB: a database of RNA-binding specificities. Nucleic acids research *39*, D301-D308

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., and Piccolo, S. (2003). Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads. Cell *113*, 301-314.

Creighton, C.J., Reid, J.G., and Gunaratne, P.H. (2009). Expression profiling of microRNAs by deep sequencing. Brief Bioinform *10*, 490-497.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R*., et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Daly, A.C., Randall, R.A., and Hill, C.S. (2008). Transforming growth factor beta-induced Smad1/5 phosphorylation in epithelial cells is mediated by novel receptor complexes and is essential for anchorage-independent growth. Molecular and Cellular Biology *28*, 6889-6902.

Davis, B.N., Hilyard, A.C., Lagna, G., and Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. Nature *454*, 56-61.

de Moor, C.H., Meijer, H., and Lissenden, S. (2005). Mechanisms of translational control by the 3' UTR in development and differentiation. Semin Cell Dev Biol *16*, 49-58.

Dickmeis, T., Aanstad, P., Clark, M., Fischer, N., Herwig, R., Mourrain, P., Blader, P., Rosa, F., Lehrach, H., and Strahle, U. (2001). Identification of Nodal signaling targets by array analysis of induced complex probes. Dev Dyn *222*, 571-580.

Ding, J., Yang, L., Yan, Y.T., Chen, A., Desai, N., Wynshaw-Boris, A., and Shen, M.M. (1998). Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. Nature *395*, 702-707.

Ding, S.W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell *130*, 413-426.

Dorey, K., and Hill, C.S. (2006). A novel Cripto-related protein reveals an essential role for EGF-CFCs in Nodal signalling in Xenopus embryos. Dev Biol *292*, 303-316.

Eimon, P.M., and Harland, R.M. (2002). Effects of heterodimerization and proteolytic processing on Derriere and Nodal activity: implications for mesoderm induction in Xenopus. Development *129*, 3089-3103.

Elinson, R.P., and Rowning, B. (1988). A transient array of parallel microtubules in frog eggs: potential tracks for a cytoplasmic rotation that specifies the dorso-ventral axis. Dev Biol *128*, 185-197.

Engleka, M.J., Craig, E.J., and Kessler, D.S. (2001). VegT activation of Sox17 at the midblastula transition alters the response to Nodal signals in the vegetal endoderm domain. Dev Biol *237*, 159-172.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol *15*, 346-353.

Farazi, T.A., Horlings, H.M., Ten Hoeve, J.J., Mihailovic, A., Halfwerk, H., Morozov, P., Brown, M., Hafner, M., Reyal, F., van Kouwenhove, M.*, et al.* (2011). MicroRNA Sequence and Expression Analysis in Breast Tumors by Deep Sequencing. Cancer Res *71*, 4443-4453.

Faure, S., Lee, M.A., Keller, T., ten Dijke, P., and Whitman, M. (2000). Endogenous patterns of TGFbeta superfamily signaling during early Xenopus development. Development *127*, 2917-2931.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S.*, et al.* (2011). Ensembl 2011. Nucleic Acids Res *39*, D800-806.

Flynt, A.S., Li, N., Thatcher, E.J., Solnica-Krezel, L., and Patton, J.G. (2007). Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. Nat Genet *39*, 259-263.

Friess, H., Yamanaka, Y., Buchler, M., Kobrin, M.S., Tahara, E., and Korc, M. (1994). Cripto, a member of the epidermal growth factor family, is over-expressed in human pancreatic cancer and chronic pancreatitis. Int J Cancer *56*, 668-674.

Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M., and Gerber, A.P. (2008). Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. PLoS ONE *3*, e3164.

Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. Proc Natl Acad Sci U S A *103*, 4487-4492.

Gerhart, J., Danilchik, M., Doniach, T., Roberts, S., Rowning, B., and Stewart, R. (1989). Cortical rotation of the Xenopus egg: consequences for the anteroposterior pattern of embryonic dorsal development. Development *107 Suppl*, 37-51.

Germain, S., Howell, M., Esslemont, G.M., and Hill, C.S. (2000). Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev *14*, 435-451.

Gessert, S., Bugner, V., Tecza, A., Pinker, M., and Kühl, M. (2010). FMR1/FXR1 and the miRNA pathway are required for eye and neural crest development. Dev Biol *341*, 222-235.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Gill, M.E., Hu, Y.C., Lin, Y., and Page, D.C. (2011). Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. Proc Natl Acad Sci U S A *108*, 7443-7448.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. Science *308*, 833-838.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Glazov, E.A., Cottee, P.A., Barris, W.C., Moore, R.J., Dalrymple, B.P., and Tizard, M.L. (2008). A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res *18*, 957-964.

Goldstrohm, A.C., Hook, B.A., Seay, D.J., and Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol *13*, 533-539.

Gonzalez, S., Klatt, P., Delgado, S., Conde, E., Lopez-Rios, F., Sanchez-Cespedes, M., Mendez, J., Antequera, F., and Serrano, M. (2006). Oncogenic activity of Cdc6 through repression of the INK4/ARF locus. Nature *440*, 702-706.

Gonzalez, S., Pisano, D.G., and Serrano, M. (2008). Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. Cell Cycle *7*, 2601-2608.

Graff, J.M. (1997). Embryonic patterning: to BMP or not to BMP, that is the question. Cell *89*, 171-174.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol *36*, 59-74.

Grande, C., and Patel, N.H. (2009). Nodal signalling is involved in left-right asymmetry in snails. Nature *457*, 1007-1011.

Gray, P.C., Shani, G., Aung, K., Kelber, J., and Vale, W. (2006). Cripto binds transforming growth factor beta (TGF-β) and inhibits TGF-β signaling. Mol Cell Biol *26*, 9268-9278.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-144.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., and Schier, A.F. (1999). The EGF-CFC protein one-eyed pinhead is essential for Nodal signaling. Cell *97*, 121-132.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Gurdon, J.B., Harger, P., Mitchell, A., and Lemaire, P. (1994). Activin signalling and response to a morphogen gradient. Nature *371*, 487-492.

Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods *44*, 3-12.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M.*, et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J *21*, 4671-4679.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hansen, C.S., Marion, C.D., Steele, K., George, S., and Smith, W.C. (1997). Direct neural induction and selective inhibition of mesoderm and epidermis inducers by Xnr3. Development *124*, 483-492.

Hardy, K.M., Kirschmann, D.A., Seftor, E.A., Margaryan, N.V., Postovit, L.M., Strizzi, L., and Hendrix, M.J. (2010). Regulation of the embryonic morphogen Nodal by Notch4 facilitates manifestation of the aggressive melanoma phenotype. Cancer Res *70*, 10340-10350.

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A., and Massague, J. (2000). OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. Cell *100*, 229-240.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L.*, et al.* (2010). The genome of the Western clawed frog Xenopus tropicalis. Science *328*, 633-636.

Hemmati-Brivanlou, A., and Thomsen, G.H. (1995). Ventral mesodermal patterning in Xenopus embryos: expression patterns and activities of BMP-2 and BMP-4. Dev Genet *17*, 78-89.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Hopwood, N.D., Pluck, A., and Gurdon, J.B. (1989). MyoD expression in the forming somites is an early response to mesoderm induction in Xenopus embryos. EMBO J *8*, 3409-3417.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B.*, et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Howell, M., and Hill, C.S. (1997). XSmad2 directly activates the activin-inducible, dorsal mesoderm gene XFKH1 in Xenopus embryos. EMBO J *16*, 7411-7421.

Howell, M., Mohun, T.J., and Hill, C.S. (2001). Xenopus Smad3 is specifically expressed in the chordoneural hinge, notochord and in the endocardium of the developing heart. Mech Dev *104*, 147-150.

Hudson, C., Clements, D., Friday, R.V., Stott, D., and Woodland, H.R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. Cell *91*, 397-405.

Huntzinger, E., Braun, J.E., Heimstadt, S., Zekri, L., and Izaurralde, E. (2010). Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. EMBO J *29*, 4146-4160.

Ibanez-Ventoso, C., Vora, M., and Driscoll, M. (2008). Sequence relationships among C. elegans, D. melanogaster and human microRNAs highlight the extensive conservation of microRNAs in biology. PLoS ONE *3*, e2818.

Jannot, G., Boisvert, M.E., Banville, I.H., and Simard, M.J. (2008). Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in C. elegans. RNA *14*, 829-835.

Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. (2006). Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. Nat Struct Mol Biol *13*, 787-792.

Johnson, C.D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., Wang, X., Shelton, J., Shingara, J.*, et al.* (2007). The let-7 microRNA represses cell proliferation pathways in human cells. Cancer Res *67*, 7713-7722.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. (2005). RAS is regulated by the let-7 microRNA family. Cell *120*, 635-647.

Jung, C.H., Hansen, M.A., Makunin, I.V., Korbie, D.J., and Mattick, J.S. (2010). Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. BMC Genomics *11*, 77.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res *110*, 462-467.

Kalmykova, A.I., Klenov, M.S., and Gvozdev, V.A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic Acids Res *33*, 2052-2059.

Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J.T., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J *29*, 3688-3700.

Kannan, S., De Santis, M., Lohmeyer, M., Riese, D.J., 2nd, Smith, G.H., Hynes, N., Seno, M., Brandt, R., Bianco, C., Persico, G.*, et al.* (1997). Cripto enhances the tyrosine phosphorylation of Shc and activates mitogen-activated protein kinase (MAPK) in mammary epithelial cells. J Biol Chem *272*, 3330-3335.

Kapsimali, M., Kloosterman, W.P., de Bruijn, E., Rosa, F., Plasterk, R.H., and Wilson, S.W. (2007). MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. Genome Biol *8*, R173.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol *5*, e57.

Kato, M., de Lencastre, A., Pincus, Z., and Slack, F.J. (2009). Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. Genome Biol *10*, R54.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep *8*, 763-769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137-1140.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Orom, U.A.*, et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A., Elkon, R., and Agami, R. (2010). A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol *12*, 1014-1020.

Khurana, J.S., Xu, J., Weng, Z., and Theurkauf, W.E. (2010). Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. PLoS Genet *6*, e1001246.

Kim, D.H., Saetrom, P., Snove, O., Jr., and Rossi, J.J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. Proc Natl Acad Sci U S A *105*, 16230-16235.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kim, Y.-K., Heo, I., and Kim, V.N. (2010). Modifications of Small RNAs and Their Associated Proteins. Cell *143*, 703-709.

Kinoshita, N., Minshull, J., and Kirschner, M.W. (1995). The identification of two novel ligands of the FGF receptor by a yeast screening method and their activity in Xenopus development. Cell *83*, 621-630.

Kiriakidou, M., Tan, G.S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P.T., and Mourelatos, Z. (2007). An mRNA m7G cap binding-like motif within human Ago2 represses translation. Cell *129*, 1141-1151.

Kirino, Y., and Mourelatos, Z. (2007). 2'-O-methyl modification in mouse piRNAs and its methylase. Nucleic Acids Symp Ser (Oxf), 417-418.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic

axis specification through activation of an ATR/Chk2 DNA damage response. Dev Cell *12*, 45-55.

Klenov, M.S., Lavrov, S.A., Stolyarenko, A.D., Ryazansky, S.S., Aravin, A.A., Tuschl, T., and Gvozdev, V.A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res *35*, 5430-5438.

Kloosterman, W.P., Wienholds, E., de Bruijn, E., Kauppinen, S., and Plasterk, R.H. (2006). In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. Nat Methods *3*, 27-29.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene *30*, 1956-1962.

Krieg, P.A., Varnum, S.M., Wormington, W.M., and Melton, D.A. (1989). The mRNA encoding elongation factor 1-alpha (EF-1 alpha) is a major transcript at the midblastula transition in Xenopus. Dev Biol *133*, 93-100.

Krützfeldt, J., Poy, M.N., and Stoffel, M. (2006). Strategies to determine the biological function of microRNAs. Nat Genet *38 Suppl*, S14-19.

Kuchenbauer, F., Morin, R.D., Argiropoulos, B., Petriv, O.I., Griffith, M., Heuser, M., Yung, E., Piper, J., Delaney, A., Prabhu, A.L.*, et al.* (2008). In-depth characterization of the microRNA transcriptome in a leukemia progression model. Genome Res *18*, 1787-1797.

Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R., and Jacks, T. (2007). Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet *39*, 673-677.

Kuniyasu, H., Yoshida, K., Yokozaki, H., Yasui, W., Ito, H., Toge, T., Ciardiello, F., Persico, M.G., Saeki, T., Salomon, D.S.*, et al.* (1991). Expression of cripto, a novel gene of the epidermal growth factor family, in human gastrointestinal carcinomas. Jpn J Cancer Res *82*, 969-973.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W.*, et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev *22*, 908-917.

Kuroda, R., Endo, B., Abe, M., and Shimizu, M. (2009). Chiral blastomere arrangement dictates zygotic left-right asymmetry pathway in snails. Nature *462*, 790-794.

Larocque, D., Galarneau, A., Liu, H.N., Scott, M., Almazan, G., and Richard, S. (2005). Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. Nat Neurosci *8*, 27-33.

Larocque, D., Pilotte, J., Chen, T., Cloutier, F., Massie, B., Pedraza, L., Couture, R., Lasko, P., Almazan, G., and Richard, S. (2002). Nuclear retention of MBP mRNAs in the quaking viable mice. Neuron *36*, 815-829.

Latinkic, B.V., and Smith, J.C. (1999). Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. Development *126*, 1769-1779.

Lau, N.C., Ohsumi, T., Borowsky, M., Kingston, R.E., and Blower, M.D. (2009). Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. EMBO J *28*, 2945-2958.

Lebedeva, S., Jens, M., Theil, K., Schwanhausser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Mol Cell *43*, 340-352.

Lee, M.H., and Schedl, T. (2001). Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for C. elegans germ cell development. Genes Dev *15*, 2408-2420.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

Levy, J.A., Virolainen, M., and Defendi, V. (1968). Human lymphoblastoid lines from lymph node and spleen. Cancer *22*, 517-524.

Li, L.C., Okino, S.T., Zhao, H., Pookot, D., Place, R.F., Urakami, S., Enokida, H., and Dahiya, R. (2006). Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci U S A *103*, 17337-17342.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463-2476.

Linsen, S.E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., de Bruijn, E., Voest, E.E.*, et al.* (2009). Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods *6*, 474-476.

Little, S.C., and Mullins, M.C. (2009). Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. Nat Cell Biol *11*, 637-643.

Liu, I.M., Schilling, S.H., Knouse, K.A., Choy, L., Derynck, R., and Wang, X.-F. (2009). TGFbeta-stimulated Smad1/5 phosphorylation requires the ALK5 L45 loop and mediates the pro-migratory TGFbeta switch. The EMBO Journal *28*, 88-98.

Liu, K., Liu, Y., Mo, W., Qiu, R., Wang, X., Wu, J.Y., and He, R. (2011). MiR-124 regulates early neurogenesis in the optic vesicle and forebrain, targeting NeuroD1. Nucleic Acids Res *39*, 2869-2879.

Liu, S., Li, D., Li, Q., Zhao, P., Xiang, Z., and Xia, Q. (2010). MicroRNAs of Bombyx mori identified by Solexa sequencing. BMC Genomics *11*, 148.

Lopez de Silanes, I., Zhan, M., Lal, A., Yang, X., and Gorospe, M. (2004). Identification of a target RNA motif for RNA-binding protein HuR. Proc Natl Acad Sci U S A *101*, 2987-2992.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A.*, et al.* (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834-838.

Lucchetta, E.M., Carthew, R.W., and Ismagilov, R.F. (2009). The endo-siRNA pathway is essential for robust development of the Drosophila embryo. PLoS ONE *4*, e7576.

Lund, E., Liu, M., Hartley, R.S., Sheets, M.D., and Dahlberg, J.E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA *15*, 2351-2363.

Lund, E., Sheets, M.D., Imboden, S.B., and Dahlberg, J.E. (2011). Limiting Ago protein restricts RNAi and microRNA biogenesis during early development in Xenopus laevis. Genes Dev *25*, 1121-1131.

Ma, L., Teruya-Feldstein, J., and Weinberg, R.A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature *449*, 682-688.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M.*, et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Maniar, J.M., and Fire, A.Z. (2011). EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. Curr Biol *21*, 449-459.

Marcet, B., Chevalier, B., Luxardi, G., Coraux, C., Zaragosi, L.E., Cibois, M., Robbe-Sermesant, K., Jolly, T., Cardinaud, B., Moreilhon, C.*, et al.* (2011). Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. Nat Cell Biol *13*, 693-699.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res *18*, 1509-1517.

Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., and Spruce, T. (2010a). A MicroRNA Targeting Dicer for Metastasis Control. Cell *141*, 1195-1207.

Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T.*, et al.* (2010b). A MicroRNA targeting dicer for metastasis control. Cell *141*, 1195-1207.

Martello, G., Zacchigna, L., Inui, M., Montagner, M., Adorno, M., Mamidi, A., Morsut, L., Soligo, S., Tran, U., Dupont, S.*, et al.* (2007). MicroRNA control of Nodal signalling. Nature *449*, 183-188.

Martens-Uzunova, E.S., Jalava, S.E., Dits, N.F., van Leenders, G.J., Moller, S., Trapman, J., Bangma, C.H., Litman, T., Visakorpi, T., and Jenster, G. (2011). Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. Oncogene.

Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S.*, et al.* (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. Science *317*, 1764-1767.

Meijer, H.A., Radford, H.E., Wilson, L.S., Lissenden, S., and de Moor, C.H. (2007). Translational control of maskin mRNA by its 3' untranslated region. Biol Cell *99*, 239-250.

Mendell, J.T. (2008). miRiad roles for the miR-17-92 cluster in development and disease. Cell *133*, 217-222.

Meno, C., Gritsman, K., Ohishi, S., Ohfuji, Y., Heckscher, E., Mochida, K., Shimono, A., Kondoh, H., Talbot, W.S., Robertson, E.J.*, et al.* (1999). Mouse Lefty2 and zebrafish antivin are feedback inhibitors of Nodal signaling during vertebrate gastrulation. Mol Cell *4*, 287-298.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature *447*, 1126-1129.

Moody, S.A. (1987). Fates of the blastomeres of the 32-cell-stage Xenopus embryo. Dev Biol *122*, 300-319.

Moon, R.T., and Kimelman, D. (1998). From cortical rotation to organizer gene expression: toward a molecular explanation of axis specification in Xenopus. Bioessays *20*, 536-545.

Moraes, K.C., Wilusz, C.J., and Wilusz, J. (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. RNA *12*, 1084-1091.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M.*, et al.* (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res *18*, 610-621.

Morris, A.R., Mukherjee, N., and Keene, J.D. (2008). Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets. Mol Cell Biol *28*, 4093-4103.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289-1292.

Mortensen, R.D., Serra, M., Steitz, J.A., and Vasudevan, S. (2011). Posttranscriptional activation of gene expression in Xenopus laevis oocytes by microRNA-protein complexes (microRNPs). Proc Natl Acad Sci U S A *108*, 8281-8286.

Moshkovich, N., and Lei, E.P. (2010). HP1 recruitment in the absence of argonaute proteins in Drosophila. PLoS Genet *6*, e1000880.

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative Regulatory Mapping Indicates that the RNA-Binding Protein HuR Couples Pre-mRNA Processing and mRNA Stability. Mol Cell *43*, 327-339.

Newman, C.S., Chia, F., and Krieg, P.A. (1997). The XHex homeobox gene is expressed during development of the vascular endothelium: overexpression leads to an increase in vascular endothelial cell number. Mech Dev *66*, 83-93.

Newport, J., and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675-686.

Newport, J., and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell *30*, 687-696.

Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. RNA *13*, 1911-1922.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89-100.

Okamura, K., Robine, N., Liu, Y., Liu, Q., and Lai, E.C. (2011). R2D2 organizes small regulatory RNA pathways in Drosophila. Mol Cell Biol *31*, 884-896.

Onuma, Y., Yeo, C.-Y., and Whitman, M. (2006a). XCR2, one of three Xenopus EGF-CFC genes, has a distinct role in the regulation of left-right patterning. Development *133*, 237-250.

Onuma, Y., Yeo, C.Y., and Whitman, M. (2006b). XCR2, one of three Xenopus EGF-CFC genes, has a distinct role in the regulation of left-right patterning. Development *133*, 237-250.

Osada, S.I., Saijoh, Y., Frisch, A., Yeo, C.Y., Adachi, H., Watanabe, M., Whitman, M., Hamada, H., and Wright, C.V. (2000). Activin/Nodal responsiveness and asymmetric expression of a Xenopus Nodal-related gene converge on a FAST-regulated module in intron 1. Development *127*, 2503-2514.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nat Protoc *3*, 1077-1084.

Pang, K.C., Stephen, S., Engstrom, P.G., Tajul-Arifin, K., Chen, W., Wahlestedt, C., Lenhard, B., Hayashizaki, Y., and Mattick, J.S. (2005). RNAdb--a comprehensive mammalian noncoding RNA database. Nucleic Acids Res *33*, D125-130.

Paris, J., Osborne, H.B., Couturier, A., Le Guellec, R., and Philippe, M. (1988). Changes in the polyadenylation of specific stable RNA during the early development of Xenopus laevis. Gene *72*, 169-176.

Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T., and De Robertis, E.M. (1999). The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. Nature *397*, 707-710.

Piccolo, S., Sasai, Y., Lu, B., and De Robertis, E.M. (1996). Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. Cell *86*, 589-598.

Prosser, H.M., Koike-Yusa, H., Cooper, J.D., Law, F.C., and Bradley, A. (2011). A resource of vectors and ES cells for targeted deletion of microRNAs in mice. Nat Biotechnol.

Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res *37*, D32-36.

Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). piRNA Production Requires Heterochromatin Formation in Drosophila. Curr Biol *21*, 1373-1379.

Rao, P.K., Kumar, R.M., Farkhondeh, M., Baskerville, S., and Lodish, H.F. (2006). Myogenic factors that regulate expression of muscle-specific microRNAs. Proceedings of the National Academy of Sciences *103*, 8721-8726.

Reissmann, E., Jornvall, H., Blokzijl, A., Andersson, O., Chang, C., Minchiotti, G., Persico, M.G., Ibanez, C.F., and Brivanlou, A.H. (2001). The orphan receptor ALK7 and the Activin receptor ALK4 mediate signaling by Nodal proteins during vertebrate development. Genes Dev *15*, 2010-2022.

Reversade, B., and De Robertis, E.M. (2005). Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. Cell *123*, 1147-1160.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincead-Beal, C., Kulkarni, P.*, et al.* (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia *9*, 166-180.

Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature *433*, 477-480.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Rosa, A., Spagnoli, F.M., and Brivanlou, A.H. (2009). The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. Dev Cell *16*, 517-527.

Ross, S., Cheung, E., Petrakis, T.G., Howell, M., Kraus, W.L., and Hill, C.S. (2006). Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. EMBO J *25*, 4490-4502.

Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128-1132.

Rowning, B.A., Wells, J., Wu, M., Gerhart, J.C., Moon, R.T., and Larabell, C.A. (1997). Microtubule-mediated transport of organelles and localization of beta-catenin to the future dorsal side of Xenopus eggs. Proc Natl Acad Sci U S A *94*, 1224-1229.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruiz-Echevarria, M.J., and Peltz, S.W. (2000). The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell *101*, 741-751.

Saijoh, Y., Adachi, H., Sakuma, R., Yeo, C.Y., Yashiro, K., Watanabe, M., Hashiguchi, H., Mochida, K., Ohishi, S., Kawabata, M.*, et al.* (2000). Left-right asymmetric expression of lefty2 and Nodal is induced by a signaling pathway that includes the transcription factor FAST2. Mol Cell *5*, 35-47.

Saloman, D.S., Bianco, C., Ebert, A.D., Khan, N.I., De Santis, M., Normanno, N., Wechselberger, C., Seno, M., Williams, K., Sanicola, M.*, et al.* (2000). The EGF-CFC family: novel epidermal growth factor-related proteins in development and cancer. Endocr Relat Cancer *7*, 199-226.

Salomon, D.S., Bianco, C., and De Santis, M. (1999). Cripto: a novel epidermal growth factor (EGF)-related peptide in mammary gland development and neoplasia. Bioessays *21*, 61-70.

Samson, M.L. (2008). Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins. BMC Genomics *9*, 392.

Sander, V., Reversade, B., and De Robertis, E.M. (2007). The opposing homeobox genes Goosecoid and Vent1/2 self-regulate Xenopus patterning. EMBO J *26*, 2955-2965.

Scharf, S.R., and Gerhart, J.C. (1983). Axis determination in eggs of Xenopus laevis: a critical period before first cleavage, identified by the common effects of cold, pressure and ultraviolet irradiation. Dev Biol *99*, 75-87.

Schier, A.F., and Shen, M.M. (2000). Nodal signalling in vertebrate development. Nature *403*, 385-389.

Schiffer, S.G., Foley, S., Kaffashan, A., Hronowski, X., Zichittella, A.E., Yeo, C.Y., Miatkowski, K., Adkins, H.B., Damon, B., Whitman, M.*, et al.* (2001). Fucosylation of Cripto is required for its ability to facilitate Nodal signaling. J Biol Chem *276*, 37769-37778.

Schroeder, M.M., and Gard, D.L. (1992). Organization and regulation of cortical microtubules during the first cell cycle of Xenopus eggs. Development *114*, 699-709.

Schumacher, B., Hanazawa, M., Lee, M.H., Nayak, S., Volkmann, K., Hofmann, E.R., Hengartner, M., Schedl, T., and Gartner, A. (2005). Translational repression of C. elegans p53 by GLD-1 regulates DNA damage-induced apoptosis. Cell *120*, 357-368.

Seitz, H., Ghildiyal, M., and Zamore, P.D. (2008). Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. Curr Biol *18*, 147-151.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.

Sexton, T., Schober, H., Fraser, P., and Gasser, S.M. (2007). Gene regulation through nuclear organization. Nat Struct Mol Biol *14*, 1049-1055.

Shi, R., and Chiang, V.L. (2005). Facile means for quantifying microRNA expression by real-time PCR. Biotechniques *39*, 519-525.

Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R., Shkumatava, A., and Bartel, D.P. (2010). Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell *38*, 789-802.

Sinner, D., Rankin, S., Lee, M., and Zorn, A.M. (2004). Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. Development *131*, 3069-3080.

Smibert, P., and Lai, E.C. (2008). Lessons from microRNA mutants in worms, flies and mice. Cell Cycle *7*, 2500-2508.

Smith, J.C., Price, B.M., Green, J.B., Weigel, D., and Herrmann, B.G. (1991). Expression of a Xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. Cell *67*, 79-87.

Spemann, H. (1938). Embryonic Development and Induction. New Haven, CT: Yale University Press.

Spemann, H., and Mangold, H. (2001). Induction of embryonic primordia by implantation of organizers from a different species. 1923. Int J Dev Biol *45*, 13-38.

Strizzi, L., Bianco, C., Normanno, N., Seno, M., Wechselberger, C., Wallace-Jones, B., Khan, N.I., Hirota, M., Sun, Y., Sanicola, M.*, et al.* (2004). Epithelial mesenchymal transition is a characteristic of hyperplasias and tumors in mammary gland from MMTV-Cripto-1 transgenic mice. J Cell Physiol *201*, 266-276.

Strizzi, L., Postovit, L.M., Margaryan, N.V., Lipavsky, A., Gadiot, J., Blank, C., Seftor, R.E., Seftor, E.A., and Hendrix, M.J. (2009). Nodal as a biomarker for melanoma progression and a new therapeutic target for clinical intervention. Expert Rev Dermatol *4*, 67-78.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J.R., Jerzmanowski, A., and Dean, C. (2007). Small RNA-mediated chromatin silencing directed to the 3' region of the Arabidopsis gene encoding the developmental regulator, FLC. Proc Natl Acad Sci U S A *104*, 3633-3638.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K.*, et al.* (2009). Tiny RNAs associated with transcription start sites in animals. Nat Genet *41*, 572-578.

Takahashi, S., Yokota, C., Takano, K., Tanegashima, K., Onuma, Y., Goto, J., and Asashima, M. (2000). Two novel Nodal-related genes initiate early inductive events in Xenopus Nieuwkoop center. Development *127*, 5319-5329.

Takeda, Y., Mishima, Y., Fujiwara, T., Sakamoto, H., and Inoue, K. (2009). DAZL relieves miRNA-mediated repression of germline mRNAs by controlling poly(A) tail length in zebrafish. PLoS ONE *4*, e7513.

Tanaka, C., Sakuma, R., Nakamura, T., Hamada, H., and Saijoh, Y. (2007). Long-range action of Nodal requires interaction with GDF1. Genes Dev *21*, 3272-3282.

Tanegashima, K., Yokota, C., Takahashi, S., and Asashima, M. (2000). Expression cloning of Xantivin, a Xenopus lefty/antivin-related gene, involved in the regulation of activin signaling during mesoderm induction. Mech Dev *99*, 3-14.

Tang, G.-Q., and Maxwell, E.S. (2008). Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. Genome Res *18*, 104-112.

Tao, Q., Yokota, C., Puck, H., Kofron, M., Birsoy, B., Yan, D., Asashima, M., Wylie, C.C., Lin, X., and Heasman, J. (2005). Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. Cell *120*, 857-871.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124-1128.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

Topczewska, J.M., Postovit, L.M., Margaryan, N.V., Sam, A., Hess, A.R., Wheaton, W.W., Nickoloff, B.J., Topczewski, J., and Hendrix, M.J. (2006). Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. Nat Med *12*, 925-932.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922-926.

Vasudevan, S., Tong, Y., and Steitz, J.A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. Science *318*, 1931-1934.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Verdel, A., Vavasseur, A., Le Gorrec, M., and Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol *53*, 245-257.

von Bubnoff, A., Peiffer, D.A., Blitz, I.L., Hayata, T., Ogata, S., Zeng, Q., Trunnell, M., and Cho, K.W. (2005). Phylogenetic footprinting and genome scanning identify vertebrate BMP response elements and new target genes. Dev Biol *281*, 210-226.

Walker, J.C., and Harland, R.M. (2008). Expression of microRNAs during embryonic development of Xenopus tropicalis. Gene Expr Patterns *8*, 452-456.

Wang, X.B., Wu, Q., Ito, T., Cillo, F., Li, W.X., Chen, X., Yu, J.L., and Ding, S.W. (2010a). RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci U S A *107*, 484-489.

Wang, Y., Lacroix, G., Haines, J., Doukhanine, E., Almazan, G., and Richard, S. (2010b). The QKI-6 RNA binding protein localizes with the MBP mRNAs in stress granules of glial cells. PLoS ONE *5*, e12824.

Watanabe, K., Meyer, M.J., Strizzi, L., Lee, J.M., Gonzales, M., Bianco, C., Nagaoka, T., Farid, S.S., Margaryan, N., Hendrix, M.J.*, et al.* (2010). Cripto-1 is a cell surface marker for a tumorigenic, undifferentiated subpopulation in human embryonal carcinoma cells. Stem Cells *28*, 1303-1314.

Watanabe, T., Takeda, A., Mise, K., Okuno, T., Suzuki, T., Minami, N., and Imai, H. (2005). Stage-specific expression of microRNAs during Xenopus development. FEBS Lett *579*, 318-324.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wechselberger, C., Strizzi, L., Kenney, N., Hirota, M., Sun, Y., Ebert, A., Orozco, O., Bianco, C., Khan, N.I., Wallace-Jones, B.*, et al.* (2005). Human Cripto-1 overexpression in the mouse mammary gland results in the development of hyperplasia and adenocarcinoma. Oncogene *24*, 4094-4105.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. RNA *12*, 256-262.

Wellner, U., Schubert, J., Burk, U.C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., zur Hausen, A.*, et al.* (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat Cell Biol *11*, 1487-1495.

Wharton, R.P., and Aggarwal, A.K. (2006). mRNA regulation by Puf domain proteins. Sci STKE *2006*, pe37.

Wharton, R.P., Sonoda, J., Lee, T., Patterson, M., and Murata, Y. (1998). The Pumilio RNA-binding domain is also a translational regulator. Mol Cell *1*, 863-872.

Wienholds, E., Kloosterman, W.P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H.R., Kauppinen, S., and Plasterk, R.H. (2005). MicroRNA expression in zebrafish embryonic development. Science *309*, 310-311.

Wilczynska, A., Minshall, N., Armisen, J., Miska, E.A., and Standart, N. (2009). Two Piwi proteins, Xiwi and Xili, are expressed in the Xenopus female germline. RNA *15*, 337-345.

Wilson, P.A., and Hemmati-Brivanlou, A. (1995). Induction of epidermis and inhibition of neural fate by Bmp-4. Nature *376*, 331-333.

Wu, M.Y., and Hill, C.S. (2009). TGF-β superfamily signaling in embryonic development and homeostasis. Developmental Cell *16*, 329-343.

Wu, M.Y., Ramel, M.C., Howell, M., and Hill, C.S. (2011). SNW1 is a critical regulator of spatial BMP activity, neural plate border formation, and neural crest specification in vertebrate embryos. PLoS Biol *9*, e1000593.

Yabe, S., Tanegashima, K., Haramoto, Y., Takahashi, S., Fujii, T., Kozuma, S., Taketani, Y., and Asashima, M. (2003). FRL-1, a member of the EGF-CFC family, is essential for neural differentiation in Xenopus early development. Development *130*, 2071-2081.

Yan, Y.T., Liu, J.J., Luo, Y., E, C., Haltiwanger, R.S., Abate-Shen, C., and Shen, M.M. (2002). Dual roles of Cripto as a ligand and coreceptor in the Nodal signaling pathway. Mol Cell Biol *22*, 4439-4449.

Yan, Z., Hu, H.Y., Jiang, X., Maierhofer, V., Neb, E., He, L., Hu, Y., Hu, H., Li, N., Chen, W.*, et al.* (2011). Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res.

Yang, X., Feng, M., Jiang, X., Wu, Z., Li, Z., Aau, M., and Yu, Q. (2009). miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb-E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. Genes Dev *23*, 2388-2393.

Yeo, C., and Whitman, M. (2001). Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. Mol Cell *7*, 949-957.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. Nature *476*, 228-231.

Yoon, H.J., Hong, J.S., Shin, W.J., Lee, Y.J., Hong, K.O., Lee, J.I., Hong, S.P., and Hong, S.D. (2011). The role of Cripto-1 in the tumorigenesis and progression of oral squamous cell carcinoma. Oral Oncol.

Zhang, J., and King, M.L. (1996). Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. Development *122*, 4119-4129.

Zhang, J., Talbot, W.S., and Schier, A.F. (1998). Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. Cell *92*, 241-251.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A *104*, 4536-4541.

Zhang, Y., Forinash, K.D., McGivern, J., Fritz, B., Dorey, K., and Sheets, M.D. (2009). Spatially restricted translation of the xCR1 mRNA in Xenopus embryos. Molecular and Cellular Biology *29*, 3791-3802.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

Zimmerman, L.B., De Jesus-Escobar, J.M., and Harland, R.M. (1996). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. Cell *86*, 599-606.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol *25*, 221-251.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.